

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 23:12:55 ; Search time 7.88961 Seconds

(without alignments)  
548.048 Million cell updates/sec

Title: US-09-816-989A-2

Perfect score: 213  
Sequence: 1 AKYAKKAKAKAKAKAKA.....AKYAKKAKAKAKAKAAYEA 45

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

PIR 78:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103.5	48.6	210	A25550	histone H1 - sea u
2	98	46.0	394	F90725	membrane spanning
3	98	46.0	394	G85576	membrane spanning
4	96	46.0	421	JY0057	tolA protein - Esc
5	94.5	44.4	239	AE1317	hypothetical prote
6	90	42.3	243	AE1689	hypothetical prote
7	90	42.3	1390	S51364	sperm tail-specific
8	89.5	42.0	211	A28100	histone H1-beta, e
9	89.5	42.0	214	G70673	probable hupB - My
10	88	41.3	899	B87553	DNA topoisomerase
11	86.5	40.6	384	B43582	outer membrane pro
12	86	40.4	199	F70742	hypothetical prote
13	86	40.4	344	S94153	mat101-1 protein -
14	86	39.4	347	E83525	tolA protein PA097
15	85	39.9	182	S61926	histone H1 homolog
16	82	38.5	248	HSURIP	histone H1, gonada
17	82	38.5	262	T35503	1-acetylglucosyl-3-p
18	82	38.5	566	T39675	asparaginyl-tRNA s
19	81.5	38.3	311	T17688	hypothetical prote
20	80.5	37.8	277	T34625	probable NUP/P60 f
21	80.5	37.8	445	T50972	probable DNA topoi
22	80.5	37.8	952	T36664	histone H1 - trout
23	80	37.6	194	HSTR1	histone H1 - rainb
24	80	37.6	206	HSTR1R	hypothetical prote
25	80	37.6	291	AB1487	histone H1-gamma,
26	79.5	37.3	217	A26721	histone H1-gamma,
27	79	37.1	62	T30977	cycr-1 protein - C
28	79	37.1	205	S19114	histone H1 - midge
29	79	37.1	244	S40436	

30	79	37.1	347	2	AB1794	probable rotamase
31	78	36.6	217	2	S28309	hypothetical prote
32	78	36.6	231	2	S53589	histone H1 - Chlam
33	78	36.6	287	2	S45662	histone H1 - tomat
34	78	36.6	309	2	G83013	polyhydroxyalkanoa
35	77.5	36.4	376	2	ACG592	tolA protein (limp
36	77	36.2	220	2	A28456	histone H1.10 - ch
37	77	36.2	226	1	S51660	histone H1-5 [vali
38	77	36.2	523	2	B38145	invariant surface
39	77	36.2	580	2	B70868	probable transiera
40	76	35.7	229	2	I51227	histone H1A - Atri
41	76	35.7	241	2	JN0748	histone H1-II - Vo
42	76	35.7	388	2	AC0138	tolA colicin impor
43	75.5	35.4	220	2	JC5954	ribosomal protein
44	75.5	35.4	924	2	T06636	hypothetical prote
45	75	35.2	212	2	A28470	histone H1 - mouse

## ALIGNMENTS

## RESULT 1

A25550 histone H1 - sea urchin (lytechinus pictus)  
C/Species: Lytechinus pictus (painted urchin)  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 23-Jul-1999

C/Accession: A25550  
R/Keywords: J.A.; Childs, G.J.

Nucleic Acids Res. 14, 821-833, 1986

A/Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus e

A/Reference number: A25550; MUID:87040778; PMID:3022245

A/Accession: A25550

A/Molecule type: DNA

A/Residues: 1-210 <XON>

A/Cross-references: GB:X04488; NID:G9616; PIDN:CAA28177.1; PID:G9617

C/Superfamily: histone H1

C/Keywords: chromosomal protein; DNA binding; nucleosome; nucleus

Query Match 48.6%; Score 103.5; DB 2; Length 210;  
Best Local Similarity 65.1%; Pred. No. 0.0052;  
Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 AKYAKKAKAKK-ACKAYKAAEAKKAYKAAEAKKAAKAAKAA 42

DB 166 AKKAAKKPAKKPAKKAKKPAKKAKKPAKKAKKPAKKAKKAA 208

## RESULT 2

F90725 membrane spanning protein TolA (imported) - Escherichia coli (strain O157:H7, substrain  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C/Accession: F90725

R/Keywords: T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.C

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and ger

A/Reference number: A59629; MUID:21156231; PMID:11258796

A/Accession: F90725

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-394 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA834197.1; PID:G13360233; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RMD 0509952

C/Genetics:

A/Genes: EC80774

Query Match 46.0%; Score 98; DB 2; Length 394;  
Best Local Similarity 65.3%; Pred. No. 0.028; 9; Indels 6; Gaps 3;  
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

QY 2 KKYAKKAKAKA---KK--AYKAAEAKKAYKAAEAKKAAKAAKAAAYEA 45



## RESULT 7

S51364

Spectrum tail-specific protein met101(2) - fruit fly (*Drosophila hydei*)C/Species: *Drosophila hydei*

C/Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 17-Mar-1999

C/Accession: S51364; S34154

R/Neese, J.; Padmanabhan, S.; Bueneemann, H.

Eur. J. Biochem. 225, 1089-1095, 1994

A/Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represses lpha-helical rods within the extremely elongated spermatozoa of *Drosophila hydei*.

A/Reference number: S51364; PMID:95045538; PMID:79577199

A/Accession: S51364

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 &lt;NEE&gt;

A/Cross-references: EMBL:X73481

R/Neese, J.; Heinlein, U.A.O.; Bueneemann, H.

submitted to the EMBL Data Library, June 1993

A/Reference number: S34153

A/Accession: S34154

A/Molecule type: DNA

A/Residues: 1-163; 'E', 164-236; 'Q', 237-254; 257-320; 'E', 321-1390 &lt;NEE&gt;

A/Cross-references: EMBL:X73481; NID:9313201; PID:9313202

C/Genetics:

A/Gene: met101(2)

A/Cross-references: Flybase:FBgn0011816

Query Match 42.3%; Score 90; DB 2; Length 1390;

Best Local Similarity 57.8%; Pred. No. 0.41; Indels 4; Gaps 2;

Matches 26; Conservative 4; Mismatches 11; Indels 4; Gaps 2;

QY 2 KKYAKKAKAEK-ARKYKAAEAKK--AAKYKAAEKAKEAA 42

DB 662 KETAEKKKCEKAKKKEAKKKEKCAEAKKKEAKKKEKCEKAA 706

## RESULT 8

A28100

histone H1-beta, embryonic - sea urchin (*Strongylocentrotus purpuratus*)C/Species: *Strongylocentrotus purpuratus* (purple urchin)

C/Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 23-Feb-1997

C/Accession: A28100

R/Jai, Z.C.; Childs, G.

Mol. Cell. Biol. 8, 1842-1844, 1988

A/Title: Characterization of the structure and transcriptional patterns of the gene encoding histone H1-beta

A/Reference number: A28100; PMID:88246461; PMID:2837660

A/Accession: A28100

A/Molecule type: DNA

A/Residues: 1-211 &lt;LAI&gt;

A/Cross-references: GB:M20314

C/Suprafamily: histone H1

C/Keywords: chromosomal protein; DNA binding; embryo; nucleosome; nucleus

Query Match 42.0%; Score 89.5; DB 2; Length 211;

Best Local Similarity 61.9%; Pred. No. 0.11; Indels 3; Gaps 2;

Matches 26; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 1 AKKYAKKAKAEKAKKAYKAAEAKKAKYKAAEKAKEAA 42

DB 139 SKTKTKKKKKAKKAKKAKKAAK-APAKKAAK 177

## RESULT 9

G70673

Probable hupB - Mycobacterium tuberculosis (strain H37Rv)

C/Species: *Mycobacterium tuberculosis*

C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C/Accession: G70673

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, S.; Skellern, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A/Reference number: A70500; PMID:98295987; PMID:9654230

A/Accession: G70673

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-214 &lt;COL&gt;

A/Cross-references: GB:283018; GB:AL123456; NID:93261671; PID:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Gene: hupB

C/Suprafamily: histone H1

Query Match 42.0%; Score 89.5; DB 2; Length 214;

Best Local Similarity 55.6%; Pred. No. 0.11; Indels 3; Gaps 1;

Matches 25; Conservative 4; Mismatches 13; Indels 3; Gaps 1;

QY 1 AKKYAKKAKAEKAKKAYKAAEAKKAKYKAAEKAKEAA 45

DB 111 AKKYAKKAPAKKATKAAKATKAPA--KKATKAPAKKATKAA 152

## RESULT 10

B67553

DNA topoisomerase I [imported] - *Caulobacter crescentus*C/Species: *Caulobacter crescentus*

C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C/Accession: B67553

R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.O.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolo

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of *Caulobacter crescentus*.

A/Reference number: A67249; PMID:21173698; PMID:11259647

A/Accession: B67553

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-899 &lt;STO&gt;

A/Cross-references: GB:AE005673; NID:913423998; PID:AAK24422.1; GSPDB:GN00148

C/Genetics:

A/Gene: CC2451

C/Suprafamily: bacterial type I DNA topoisomerase

Query Match 41.3%; Score 88; DB 2; Length 899;

Best Local Similarity 58.7%; Pred. No. 0.45; Indels 2; Gaps 2;

Matches 27; Conservative 5; Mismatches 12; Indels 2; Gaps 2;

QY 1 AKKYAKKAKAEKAKKAYKAAEAKKAKYKAAEKAKEAA 45

DB 836 AKKAKKAAATKSKAKKESDPAKKTAA-KKPAKKKPAKKA 880

## RESULT 11

B43592

outer membrane protein TmpB - *Treponema phagedenis*C/Species: *Treponema phagedenis*

C/Date: 30-Jan-1993 #sequence\_revision 12-Mar-1993 #text\_change 08-Oct-1999

C/Accession: B43592

R/Yellon, D.B.; Limberger, R.J.; Curci, K.; Malinovsky-Rummenell, F.; Sliwinski, L.; Scho

Infect. Immun. 59, 3685-3693, 1991

A/Title: *Treponema phagedenis* encodes and expresses homologs of the *Treponema pallidum*

A/Reference number: A43592; PMID:91372983; PMID:1894368

A/Accession: B43592

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-384 &lt;YEL&gt;

A/Cross-references: GB:M58563; NID:9155066; PID:AAA27480.1; PID:9155067

A/Note: the authors translated the codon TTC for residue 316 as Tyr, and CGA for residu

C/Keywords: membrane protein

Query Match 40.6%; Score 86.5; DB 2; Length 384;

Best Local Similarity 54.5%; Pred. No. 0.32; Indels 1; Gaps 1;

Matches 24; Conservative 6; Mismatches 13; Indels 1; Gaps 1;





US-09-405-143A-1

```

; APPLICANT: Yeda Research and Development Co., Ltd.
; TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
; FILE REFERENCE: 60807-A

```

```

? CURRENT APPLICATION NUMBER: US/09/405,743A
? CURRENT FILING DATE: 1999-09-24
? NUMBER OF SEQ ID NOS: 7
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 1
? LENGTH: 35
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
? OTHER INFORMATION: PEPTIDE
US-09-405-743A-1

```

Query Match	50.7%	Score 108; DB 4;	Length 35;
Best Local Similarity	64.4%	Pred. No. 1.7e-05;	
Matches	29; Conservative	1; Mismatches	5; Indels 10; Gaps 2;

```
QY      1 AKKYAKKAKEKAKKAYAAEKKAKYEEKAASAKAAAEAAAYEA   45
          ||||| | | | | | | | | | | : ||||| |||
Db      1 AKKTAKEKA--AKKAY-----KEAKAKAAEAAPAAEAAYEA   35
```

RESULT 8  
 US-09-489-039A-13565  
 Sequence 13565, Application US/09489039A  
 Patent No. 6610836  
 GENERAL INFORMATION:  
 APPLICANT: Gary Bleton et. al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 2709.2004001  
 CURRENT FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: US/09/489,039A  
 PRIOR FILING DATE: 1999-01-29  
 NUMBER OF SEQ ID NOS: 14342  
 SEQ ID NO 13565  
 LENGTH: 469  
 TYPE: PR1  
 ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-13565

Query Match Similarity	44.6%	Score 95	DB 4	Length 469
Best Local Similarity	52.9%	Pred. No. 0.006		
Matches 27	Conservative 4	Mismatches 14	Indels 6	Gaps 1

```

QY      1 AKTAAKKAKEKAKKAYKAAEAK-----KAKYKKAATAEKAAAEAYEA 45
      | : | | : | | | : | | | | | | | | | | | | | | | |
Db      242 AAEEQKKAEEAAAKKAAQGEAEKKAGGEAAKGAAEKAAAEKAAEKAQA 292

```

RESULT 9  
 US-09-095-855-201  
 Sequence 201, Application US/09095855  
 Patent No. 6160093  
 GENERAL INFORMATION:  
 APPLICANT: Tan, Paul  
 APPLICANT: Visser, Elizabeth  
 APPLICANT: Skinner, Margot  
 APPLICANT: Preeltje, Ross  
 TITLE OF INVENTION: Compounds and Methods for  
 TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections  
 NUMBER OF SEQUENCES: 208  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Law Offices of Ann W. Speckman  
 STREET: 2601 Elliott Avenue, Suite 4185  
 CITY: Seattle, WA 98105  
 COUNTRY: US

```

; STATE: WA
; COUNTRY: USA
; ZIP: 98121
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;
; COMPUTER: IBM Compatible

```

```

? OPERATING SYSTEM: DOS
? SOFTWARE: FastSEO for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/095.855
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/705,347
? FILING DATE: 29-AUG-1996
? APPLICATION NUMBER: 08/873,970
? FILING DATE: 12-JUN-1997
? APPLICATION NUMBER: 08/997,362
? FILING DATE: 23-DEC-1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Sleath, Janet
? REGISTRATION NUMBER: 37,007
? REFERENCE/DOCKET NUMBER: 11000.1002c3
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 206-269-0565
? TELEFAX: 206-269-0563
? TELEX:
? INFORMATION FOR SEQ ID NO: 201:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 223 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
US-09-095-855-201

```

```
Query Match      42.7% Score 91 DB 3 Length 223;
Best Local Similarity 59.6%; P-Id. No. 0.0074;
Matches    28; Conservative   5; Mismatches   12; Indels   2; Gaps     2;
```

```

RESULT 10
US-09-205-426-201
; Sequence 201, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; FILE REFERENCE: 11000.1002c4
; CURRENT APPLICATION NUMBER: US/09/205,426

```

```

1  EARLIER APPLICATION NUMBER: 09/095,855
2  EARLIER FILING DATE: 1998-06-11
3  EARLIER APPLICATION NUMBER: 08/997,362
4  EARLIER FILING DATE: 1997-12-23
5  EARLIER APPLICATION NUMBER: 08/873,970
6  EARLIER FILING DATE: 1997-06-12
7  EARLIER APPLICATION NUMBER: 08/705,347
8  EARLIER FILING DATE: 1996-08-29
9  NUMBER OF SEQ IDS: 208
10 SOFTWARE: FastSeq for Windows Version 3.0
11 SEQ ID NO 201
12 LENGTH: 223
13 TYPE: PRF
14 ORGANISM: Mycobacterium vaccae
15 US-09-205-426-201

```

```

Query Match      42.7%  Score 91;  DB 4;  Length 223;
Best Local Similarity 59.6%  Pred. No. 0.0074;
Matches 28;  Conservative 5;  Mismatches 12;  Indels 2;  Gaps 2;

Qy      1  AKKVA-KKAAVEKKAKYKAAAEKAAKAYEKAA-AEKAAAEKAAEYA 45
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```



```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

```

```

Query Match          38.3%; Score 81.5; DB 4; Length 497;
Best Local Similarity 47.1%; Pred. No. 0.17;
Matches 24; Conservative 7; Mismatches 13; Indels 7; Gaps 1;

```

```

QY      2 KKVAKKAKAK-----AKKAYKAAEKKAKYERAAKAAKAAKAAKAA 45
DB      236 KKEAKELAEQARQRAAKAEQAAAQAAQAAQAAQAAQAAKATKAAKAA 286

```

```

RESULT 15
US-09-252-991A-32957
; Sequence 32957, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32957
; LENGTH: 316
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32957

```

```

Query Match          36.6%; Score 78; DB 4; Length 316;
Best Local Similarity 54.8%; Pred. No. 0.26;
Matches 23; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

```

QY      1 AKKYAKKAKAEKAKYKAAEAKKAKCYEKAAEKAKEAA 42
DB      157 AKAAKPAKPAKPAKPAKTAAPAKPAKPAKPAKPAKPAKPA 198

```

Search completed: April 20, 2004, 23:19:07  
 Job time : 9.35065 secs



RESULT 2	TOLa_ECOLI	STANDARD;	PRT;	421 AA.
ID	TOLa_ECOLI			
AC	P19934			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	TOLa protein.			
GN	TOLa OR C1M OR EXCC OR LKY OR B0739.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI TaxID=562;			
RY	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=JM105;			
RX	MEDLINE=90078104; PubMed=2687247;			
RY	Levengood S.K., Webster R.E.;			
RT	"Nucleotide sequences of the <i>tolA</i> and <i>tolB</i> genes and localization of			
RT	their products, components of a multistep translocation system in			
RT	<i>Escherichia coli</i> ."			
RL	J. Bacteriol. 171:6600-6609(1989).			
RY	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KL12 / MG1655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RY	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RY	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RY	Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,			
RY	Mau B., Shao Y.;			
RY	"The complete genome sequence of <i>Escherichia coli</i> K-12.";			
RL	Science 277:1453-1474(1997).			
RY	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KL12;			
RX	MEDLINE=97061202; PubMed=8905532;			
RY	Oshina T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,			
RY	Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kasinouchi K.,			
RY	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizubuchi K.,			
RY	Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,			
RY	Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,			
RY	Yano M., Horiuchi T.;			
RY	"A 718-bp DNA sequence of the <i>Escherichia coli</i> K-12 genome			
RT	corresponding to the 12.7-28.0 min region on the linkage map."			
RL	DNA Res. 3:137-155(1996).			
RY	[4]			
RP	DOMAINS.			
RX	MEDLINE=91296736; PubMed=2068069;			
RY	Levengood S.K., Beyer W.F. Jr., Webster R.E.;			
RY	" <i>tolA</i> , a membrane protein involved in colicin uptake contains an			
RY	extended helical region."			
RY	Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).			
RY	[5]			
RP	INTERACTION WITH PORINS.			
RX	MEDLINE=97133271; PubMed=8978668;			
RY	Deronche R., Gavioli M., Bendetelli H., Pilirov A., Lazdunski C.,			
RY	Iloddes R.;			
RY	" <i>TolA</i> cellular domain interacts with <i>Escherichia coli</i> porins."			
RL	EMBO J. 15:6408-6415(1996).			
RY	[6]			
RP	X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.			
RX	MEDLINE=99332679; PubMed=10404600;			
RY	Lubkowski J., Henneke F., Plueckthun A., Wlodawer A.;			
RY	"Filamentous phage infection: crystal structure of g3p in complex			
RY	with its coreceptor, the C-terminal domain of <i>TolA</i> ."			
RL	Structure 7:711-722(1999).			
CC	-1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A			
CC	COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE			
CC	COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL			
CC	BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION			
CC	OF BACTERIOPHAGE DNA.			
CC	-1- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPc, OMPc, PHOE			
CC	AND LAMB.			

```
CC CC -I SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC CC -----
CC DR EMBL; M28232; AAA24683.1; -.
CC DR EMBL; AE000177; AAC39833.1; -.
CC DR EMBL; D90713; BAA35405.1; -.
CC PIR; JY0057; JY0057.
CC DB; JTOI; 20-MAY-99.
CC Ecogene; EG11007; tola.
CC Transport; Protein transport; Bacteriocin transport; Transmembrane;
CC Repeat; Inner membrane; 3D-structure; Complete proteome.
CC KW DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 14 34 POTENTIAL.
CC FT DOMAIN 35 421 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 48 310 DOMAIN II (ALPHA-HELICAL).
CC FT DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
CC FT DOMAIN 224 278 10 X INDEXED REPEATS OF [ED]-K(1,2)-A(2,4).
CC FT DISULFID 363 388
CC FT HELIX 335 349
CC FT TURN 350 351
CC FT TURN 353 354
CC FT HELIX 355 358
CC FT TURN 359 360
CC FT STRAND 363 369
CC FT TURN 371 372
CC FT STRAND 375 383
CC FT HELIX 385 397
CC FT HELIX 405 412
CC FT TURN 413 414
CC FT STRAND 416 421
CC SQ SEQUENCE 421 AA; 43156 MW; 8B2FE2BA897C655E CRC64;
Oy 2 KCTAKKKAKEA---KK-AVKAAEKAKTEKAAEKAAEKAAEAAYA 45
Db 247 KKAEEKAAEKAAADKKAEEKAAADKKA--AKAAEKAAAAAKAAEA 294
Query Match 46.0%; Score 98; DB 1; Length 421;
Best Local Similarity 65.3%; Pred. No. 0.0068;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;
RESULT 3
ID MST2_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acanememe-associated protein mat101(2).
GN Mat101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=9504538; PubMed=7957199;
RA "Neesen U., Padmanabhan S., Buessmann H.;
RT "famendly arranged repeats of a novel highly charged 16-amino-acid
RT acaneme-associated protein family Dmmer101 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei.";
RL Eur. J. Biochem. 225:1089-1095 (1994).
```

```

CC -1- FUNCTION: Possible structural role in the sperm tail.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Testis. Primary spermatocytes and early
CC spermatids.
CC -1- DOMAIN: The predominant structure is alpha-helical.
CC -1- POLYMORPHISM: Length polymorphisms exist between different
CC strains, most likely caused by length variations within the tandem
CC repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73481; CAAS1876.1; -.
DR PIR; S51364; S51364.
DR FlyBase; FBgn020733; Dhyd\mesl101(2).
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007288; P:axosome assembly; IEA.
KM Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1268
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F4878 CRC64;

Query Match
Best Local Similarity 42.3%; Score 90; DB 1; Length 1391;
Matches 26; Conservative 4; Mismatches 11; Indels 4; Gaps 2;

QY 2 KKYAKKAKAEK-AKKAYKAAEAKK---AAKYKAAKAAKAAKAA 42
663 KETKEXKXCKEAKKAKKAAKKAKEKKAKEKKAKEKKAKEKKA 707

RESULT 4
DB MYCSM STANDARD; PRT; 208 AA.
AC Q9ZHC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).
DE HUP OR HLP.
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / mc(2)155;
RX MEDLINE=99110209; PubMed=8994918;
RA Lee B.H., Munagasu-Oel B., Dick T.;
RT "upregulation of a histone-like protein in dormant Mycobacterium
RT smegmatis.";
RL Mol. Genet. 260:475-479(1998).
CC -1- FUNCTION: Histone-like DNA-binding protein which is capable of
CC wrapping DNA to facilitate it, and thus to prevent its denaturation
CC under extreme environmental conditions (By similarity).
CC -1- SIMILARITY: Belongs to the bacterial histone-like protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF068138; AADI3809.1; -.
DR HSSP; P02346; LHU.

```

```

DR InterPro; IPR000119; Bac DNAbind.
DR Pfam; PF00216; Bac DNAbind; 1.
DR PRINTS; PR01727; DNABINDINGHU.
DR ProDom; PD000945; Bac DNAbind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE LIKE; 1.
KM DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90
FT DOMAIN 101 205
SQ SEQUENCE 208 AA; 21230 MW; CAF577F61F7EF09 CRC64;

Query Match
Best Local Similarity 42.0%; Score 89.5; DB 1; Length 208;
Matches 29; Conservative 3; Mismatches 13; Indels 9; Gaps 2;

QY 1 AKKYAKKAKAEK-----KKAYKAAKAAKAAK--EKAAAEKAAKAAKAA 45
111 AKKAAKAPAKKAAKATATATAAKKAPAKKAAKATAPAKKAAKATAPAKKAAKATKA 164

RESULT 5
DB MYCSM STANDARD; PRT; 211 AA.
AC P15869;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-beta, late embryonic.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Echinoidea; Echinacea; Echinacea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88246461; PubMed=2837660;
RA Lai Z.-C., Childs G.;
RT "Characterization of the structure and transcriptional patterns of
RT the gene encoding the late histone subtyp H1-beta of the sea urchin
RT Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844(1988).
CC -1- FUNCTION: Histones H1 are necessary for the condensation of
CC nucleosome chains into higher order structures.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the histone H1/5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20314; AAA30052.1; -.
DR PIR; A28100; A28100.
DR HSSP; P02359; IHST.
DR InterPro; IPR005818; Histone H1/5.
DR InterPro; IPR005819; Histone H5.
DR InterPro; IPR003216; LinkerHist N.
DR Pfam; PF00538; Linker histone; I.
DR PRINTS; PR00624; HISTONH5.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
KM Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334B8E7A CRC64;

Query Match
Best Local Similarity 42.0%; Score 89.5; DB 1; Length 211;
Matches 26; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

QY 1 AKKYAKKAKAEKAKKAYKAAEAKKAAKAAKAAKAAKAAKAAKAA 42

```



Db 139 SKKTKKVKKPAKAKKPA-AKKA--KPAKPAKKA 177

## RESULT 6

DBH\_MYCTO STANDARD; PRT; 214 AA.

AC 09XB18; 0955J5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DE DNA-binding protein HU homolog (Histone-like protein) (Hlp)  
 GN HLP OR HLP OR HUPB OR MDP1 OR MB3010C.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ANS;  
 RA Prabhakar S., Tyagi J.S., Prasad H.K.;  
 RT "HLP-A target for differentiation of M. tuberculosis and M. bovis."  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BOG / Tokyo;  
 RA Matsumoto S., Yukioka H., Matsuo T., Yamada T.;  
 RT "Identification of a novel protein generating bacterial slow growth  
 from Mycobacterium."  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglsmeier K., Camus J.-C., Medina N., Mancoor H.,  
 RA Pryor M., Dutfoy S., Grogdini S., Lacroix C., Monsenpe C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Heston R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -1- FUNCTION: Histone-like DNA-binding protein which is capable of  
 wrapping DNA to stabilize it, and thus to prevent its denaturation  
 under extreme environmental conditions (by similarity).  
 CC -1- SIMILARITY: Belongs to the bacterial histone-like protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Y18421; CAB6493.1; -  
 DR EMBL: AB013441; BA78330.1; -  
 DR EMBL: BX248344; CAD96697.1; -  
 DR HSSP: P02346; 1HUU.  
 DR InterPro: IPR000119; Bac DNABind.  
 DR Pfam: PF00216; Bac DNA binding; 1.  
 DR PRINTS: PR01727; DNABINDINGHU.  
 DR ProDom: PD000945; Bac DNABind; 1.  
 DR SMART: SM00411; BHL; 1.  
 DR PROSITE: PS00045; HISTONE\_LIKE; 1.  
 DR DNA-binding; DNA condensation; Repeat; Complete proteome.  
 FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
 FT DOMAIN 101 214 DEGENERATE REPEATS REGION.  
 FT CONFLICT 137 145 MISSING (IN REF. 1 AND 2).  
 FT CONFLICT 208 208 T -> A (IN REF. 1).  
 SO SEQUENCE 214 AA; 22187 MW; C809AF20F35544 CRC64;

Query Match 42.0%; Score 89.5; DB 1; Length 214;  
 Best Local Similarity 55.6%; Pred. No. 0.025; Mismatches 13; Indels 3; Gaps 1;  
 Matches 25; Conservative 4;

Cy 1 AKYAKAKAKKAKKAKYKAAEAKKAKTERKAAEKAABAEVA 45  
 Db 111 AKYAKAKAKKAKKAKKAKKATKAPA--RKATKAPAKKAKATKA 152

## RESULT 7

DBH\_MYCTO STANDARD; PRT; 214 AA.

AC P95109;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE 15-OCT-2003 (Rel. 42, Last annotation update)  
 DE DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa  
 laminin-2-binding protein).  
 GN HLP OR HLP OR HBP21 OR RV2986C OR MT3064 OR MTCY349.01.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglsmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Baeham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean U., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=2206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,  
 RA Beshel W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Bacteriol. 184:5479-5490(2002).

CC -1- FUNCTION: Histone-like DNA-binding protein which is capable of  
 CC wrapping DNA to stabilize it, and thus to prevent its denaturation  
 CC under extreme environmental conditions (by similarity).  
 CC -1- SIMILARITY: Belongs to the bacterial histone-like protein family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z83018; CAB05427.1; -  
 DR EMBL: AE007127; AAK47393.1; -  
 DR PIR: G70673; G70673.  
 DR HSSP: P02346; 1HUU.  
 DR TIGR: MT3064; -  
 DR TubercuList; RV2986C; -  
 DR InterPro: IPR000119; Bac DNABind.  
 DR Pfam; PF00216; Bac DNA binding; 1.



```

QY      1 AKKAKKKAAEKKAAYKAAREKKAKEKAAAEEKAAGAAYE 44
        |||::|||::|||::|||::|||::|||::|||::|||::|||
DB     178 AKEKAEEKKANDKAAREKKAKEKAK-KDRAAKEGAARAE 220

RESULT 10
RS16_COREF          STANDARD;       PRT;    168 AA.
ID   RSf30;
AC   Q11142; OSF5733;              STANDARD;       PRT;    198 AA.
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Heparin-binding hemagglutinin (Adhesin).
GN   HBHA OR RVD0475 OR MTO493 OR MRCY20G9.01 OR MB0485.
OS   Mycobacterium tuberculosis, and
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC   Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
CX   NCBI_TaxId=152794;
[1]
SEQUENCE FROM N.A.
STRAIN=Y-S-314 / AJ 12310 / DSM 44549 / JCM 11169;
RX   MEDLINE=22723752; PubMed=12840036;
RA   Nhilio Y., Nakamura Y., Kawarayashi Y., Usuda Y., Kimura E.,
RA   Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
PA   Gajdosi T.;
KT   "Comparative complete genome sequence analysis of the amino acid
RT   replacements responsible for the thermostability of Corynebacterium
RL   efficiens."/
RR   Genome Ref.: 13:1572-1579(2003).
CC   -I SIMILARITY: Belongs to the SlEP family of ribosomal proteins.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and thus statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CO   or send an email to licenses@isb-sib.ch).
-----
DR   ENBAI, AP005220; BACT18770.1; -.
DR   HAIVAP, MF_00385;-; I.
DR   InterPro; IPRO00307; Ribosomal_S16.
DR   Pfam; PF00886; Ribosomal_S16; 1.
DR   ProDom; PD003791; Ribosomal_S16; 1.
DR   TRIFRAMS; TRIGR00002; S16; 1_-
DR   PROSITE; PS00732; RIBOSOMAL_S16; FAUSE_NEG.
KW   Ribosomal protein; Complete_proteome; PAUSE_NEG.
SQ   SEQUENCE 168 AA; 18228 MM; 443CB68AA232C01 CQC64;

Query Match            40.4%; Score 86; DB 1; Length 169;
Best Local Similarity   65.7%; Pred.No. 0.045;
Matches 23; Conservative 3; Mismatches 7; Indels 2; Gaps 1,

CY      7 KAQAERAKKAYKAAREKKAATCEKKAARKAARAEN 41
        ::|||||::|||:|||::::|
DB     120 EAITEKKKAREKEKAENA--EKRAAEKAAAAEA 152

RESULT 11
HBHA_MYCTU         STANDARD;
ID   HBHA_MYCTU           STANDARD;       PRT;    198 AA.
AC   Q11142; OSF5733;              STANDARD;       PRT;    198 AA.
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Heparin-binding hemagglutinin (Adhesin).
GN   HBHA OR RVD0475 OR MTO493 OR MRCY20G9.01 OR MB0485.
OS   Mycobacterium tuberculosis, and
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CC   Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
CX   NCBI_Taxid=1773, 1765;
```

RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RC SPECIES=M.tuberculosis, and M.bovis;  
 RC STRAIN=H37Rv, ATCC 201 / H37Ra, and BCG / Paris 1173 P2;  
 RX MEDLINE=98445421, PubMed=9770536;  
 RX Menozzi F.D., Bischoff R., Fort E., Brennan M.J., Locht C.;  
 RT "Molecular characterization of the Mycobacterium heparin-binding  
 RT hemagglutinin, a mycobacterial adhesin";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:12625-12630(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;  
 RC MEDLINE=98255987, PubMed=9634220;  
 RA Cole S.T., Broese R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gae S., Barry C.B. III, Tekala F.,  
 RA Baddock K., Baehum D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagsle K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,  
 RA Stulson J.E., Taylor K., Whitehead S., Barrall B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence";  
 RL Nature 393:537-544(1998).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oehkosh;  
 RC MEDLINE=22206494, PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn J.A., Hatt D., Hickey E.,  
 RA Kolony J.F., Nelson W.C., Ormawam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utechtack T., Weidman J., Kouri H., Gill U., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL J. Bacteriol. 184:5479-5490(2002).  
 [4]  
 RP SEQUENCE OF 1-16, AND CHARACTERIZATION.  
 RC SPECIES=M.tuberculosis, and M.bovis;  
 RC STRAIN=ATCC 201 / H37Ra, and BCG / Paris 1173 P2;  
 RX MEDLINE=97188915, PubMed=9064359;  
 RX Menozzi F.D., Rouse J.H., Alavi M., Laude-Sharp M., Muller J.,  
 RA Bischoff R., Brennan M.J., Locht C.;  
 RT "Identification of a heparin-binding hemagglutinin present in  
 RT Mycobacteria";  
 RL J. Exp. Med. 184:993-1001(1996).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=AF2122/97;  
 RC MEDLINE=22709107, PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsme C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RA "The complete genome sequence of Mycobacterium bovis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 [6]  
 RP FUNCTION.  
 RC SPECIES=M.tuberculosis, and M.bovis;  
 RC STRAIN=103 and BCG;  
 RX MEDLINE=21342355, PubMed=1149276;  
 RX Pehe K., Alonso S., Blet F., Delogu G., Brennan M.J., Locht C.,  
 RA Menozzi F.D.;  
 RT "The heparin-binding haemagglutinin of M. tuberculosis is required for  
 RT extrapulmonary dissemination";  
 RL Nature 412:190-194(2001).  
 -1- FUNCTION: REQUIRED FOR EXTRA-PULMONARY DISSEMINATION. MEDIATES  
 ADHERENCE TO EPITHELIAL CELLS BY BINDING TO SULFATED  
 GLYCOSAMINOGLYCAN PRESENT AT THE SURFACE OF THESE CELLS; BINDS  
 HEPARIN, DEXTAN SULFATE, FUCIDAN AND CHONDROITIN SULFATE.  
 PROMOTES HEMAGGLUTINATION OF ERYTHROCYTES OF CERTAIN HOST SPECIES.  
 INDUCES MYCOBACTERIAL AGGREGATION.  
 -1- SUBCELLULAR LOCATION: Surface-associated.

```

CC -1- DOMAIN: HEPARIN BINDING SEEMS TO REQUIRE THE C-TERMINAL
CC DOMAIN OF HBHA. PROGRESSIVE TRUNCATIONS FROM THE C-TERMINAL END
CC DIMINISH THE AFFINITY FOR HEPARIN.
CC -1- PM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
CC PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
CC SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
CC TERMINAL DOMAIN OF HBHA.
CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
CC RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CC CONTAIN ANY.
CC -1- SIMILARITY: STRONG, TO M. LEPRAE HBHA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF074390; AAC26052.1; -
DR EMBL; Z77162; CAB00936.1; -
DR EMBL; AE006951; AAK44716.1; ALT_INIT.
DR EMBL; BX248335; CAD93348.1; -
DR PIR; F70742; F70742.
DR TIGR; MT0493; -
DR TubercuList; Rv0475; -
KM Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
KM Virulence; Complete proteome.
FT INIT MET 0
FT DOMAIN 151 193 ALA/LYS-RICH.
FT CONFLICT 120 120 R -> P (IN REF. 1).
SQ SEQUENCE 198 AA; 21403 MM; 513760F6F1EB6042 CRC64;

Query Match 40.4%; Score 86; DB 1; Length 198;
Best Local Similarity 65.6%; Pred. No. 0.052;
Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 11 EKAKKAYKAAEKAKYKAAEKAKKAA 42
Db 160 KKAAPAKKAPAKKAPAKKAPAKKAPAKKAA 191

RESULT 12
MS11 DROHY STANDARD; PRT; 344 AA.
AC 008695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Axoneme-associated protein met101(1).
GN MS1101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7224;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen U., Benemann H., Heinlein U.A.O.;
RA "The Drosophila hydei gene Dms101(1) encodes a testis-specific,
RA repetitive, axoneme-associated protein with differential abundance in
RA Y chromosome deletion mutant flies.";
RL Dev. Biol. 162:414-425(1994).
CC -1- FUNCTION: Possible structural role in the sperm tail. It is
CC associated with axonemal structures.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: Testis. Located in spermatocytes and
CC spermatid bundles.
CC -1- DOMAIN: The predominant structure is alpha-helical.

```

```

CC -1- POLYMORPHISM: The number of repeats varies between strains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73480; CAA51875.1; -
DR PIR; S34153; S34153.
DR FlyBase; FBgn0011816; Dhyd\met101(1).
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0001198; F:structural molecule activity; IEP.
DR GO; GO:0007286; P:axoneme assembly; IEP.
KM Sperm; Repeat; Multigene family.
FT DOMAIN 58 337
FT 19 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.
SQ SEQUENCE 344 AA; 37793 MM; 24C65D2510387E2A CRC64;

Query Match 40.4%; Score 86; DB 1; Length 344;
Best Local Similarity 57.8%; Pred. No. 0.084;
Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;

QY 2 KKYAKKAK-AAEKAKYKAAKAKK--AAKYEKAAEKAKKAA 42
Db 69 KEAAKKKCAKAAKKEKAAKKEKAAKKEKAAKKEKAAKCAEAA 113

RESULT 13
TOLA PSEAB STANDARD; PRT; 347 AA.
AC P50600;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tola protein.
GN TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RX MEDLINE=97113525; PubMed=8955385;
RA Denis J.J., Lafontaine E.R., Sokol P.A.;
RA "Identification and characterization of the tolQRA genes of
RA Pseudomonas aeruginosa.";
RL J. Bacteriol. 178:7059-7068(1996).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Duan K., Sokol P.A.;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RC MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Britkman P.S.L., Huftnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody U.L., Coulter S.N., Folger K.R., Kae A., Lartigat K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RA opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- FUNCTION: Involved in the tonB-independent uptake of proteins (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential).

```





DT 01-OCT-2003 (TREMBLrel. 25, last annotation update  
DE TOLA protein.  
GN TOLA OR C0818.  
OS Escherichia coli O6.



OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 NC NCB1\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=06:HI / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Weich R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Moxley H.L.T., Domeneberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:117020-117024(2002).  
 DR EMBL; AB016757; AAN79291.1; -  
 KM Complete proteome.  
 SQ SEQUENCE 421 AA; 43184 MM; DB296626F056D385 CRC64;  
 Query Match 46.0%; Score 98; DB 16; Length 421;  
 Best Local Similarity 65.3%; Pred. No. 0.051;  
 Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 45  
 DB 247 KKAEEKAAEKAAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAEA 294  
 RESULT 6  
 ID 0835A1 PRELIMINARY; PRT; 413 AA.  
 AC 0835A1;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Membrane spanning protein, required for outer membrane integrity.  
 GN TOXA OR SP0558 OR S0571.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OC NCB1\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Zhang F., Zhang X., Yang J., Yang H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 through comparison with genomes of Escherichia coli K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;  
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G., III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 flexneri serotype 2a strain 2457T."  
 RL Infect. Immun. 71:2775-2786(2003).  
 DR EMBL; AB015086; AAN42202.1; -  
 DR EMBL; AB016797; AAP16075.1; -  
 KM Complete proteome.  
 SQ SEQUENCE 413 AA; 42355 MM; 93E10F2C5DE60DE8 CRC64;  
 Query Match 44.6%; Score 95; DB 16; Length 413;  
 Best Local Similarity 63.3%; Pred. No. 0.059;  
 Matches 31; Conservative 3; Mismatches 9; Indels 6; Gaps 3;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 45

DB 239 KKAEEKAAADKKAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAEA 286  
 RESULT 7  
 ID 0815W4 PRELIMINARY; PRT; 239 AA.  
 AC 0815W4;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Hypothetical protein lmo1941.  
 GN LMO1941.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OC NCB1\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaeser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,  
 RA Charbit A., Cheuani F., Couve E., de Daruvar A., Deboux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Desauter O.,  
 RA Eutlian K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kretz U., Kuhn M., Kunz F., Kuzepkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Pucell R.,  
 RA Remmel B., Rose M., Schuener T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland U., Cossart P.;  
 RT "Comparative genomics of Listeria species."  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591981; CAD00019.1; -  
 DR PIR; AB1317; AE1317.  
 DR Listerlist; LMO1941; -  
 DR GO; GO:0016998; Pcell wall catabolism; IEA.  
 DR InterPro; IPR002482; LysM.  
 DR Pfam; PF01476; LysM; 1.  
 DR SMART; SM00257; LysM; 1.  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 239 AA; 25836 MM; 72E59D57E0D7832 CRC64;  
 Query Match 44.4%; Score 94.5; DB 16; Length 239;  
 Best Local Similarity 56.2%; Pred. No. 0.066;  
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
 QY 2 KKYAKKAAEKAA--KK--AYKAAEKAAKYEKAAEKAAEKAAEAYEA 44  
 DB 124 KKAEEKAAADKKAADKKAEEKAAADKKA--AKAAEKAAAKAAEAAEA 171  
 RESULT 8  
 ID 081H69 PRELIMINARY; PRT; 248 AA.  
 AC 081H69;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE TcdA-E operon negative regulator.  
 GN BC0954.  
 OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OC NCB1\_TaxID=226900;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22608415; PubMed=12721630;  
 RX Ivanova N., Sorokin A., Anderson I., Galleron N., Candolon B.,  
 RA Kapatal V., Bhattacharya A., Resnik G., Mikhailova N., Lapidis A.,  
 RA Chu L., Masur M., Goldsman E., Larsen N., D'Souza M., Waltons T.,  
 RA Grechkin Y., Fusch G., Hasele Korn R., Forstner M., Ehrlich S.D.,  
 RA Overbeek R., Kyrides N.;  
 RT "Genome sequence of Bacillus cereus and comparative analysis with



R	Huang C-H., Kloser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz E., Rajandream M.A., Rutherford G., Rutter S., Taylor K., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Jaylor K., Warren T., Wietorrek A., Woodward U., Barrell B.G., Parkhill J., Hopwood D.A.
RA	"Complete genome sequence of the model actinomycete Streptomyces coelicolor AJ(2).";
RT	Nature 417:141-147(2002).
RL	EMBL; AL931910; CAAC8545.1;
DR	Complete proteome.
KM	SEQUENCE 24 AA; 25524 MW; 61999D6C2A3A7B0 CRC64;
SO	SEQUENCE
Dy	Query Match 42.7%; Score 91; DB 16; Length 244; Best Local Similarity 59.5%; Pred. No. 0.13; Indels 2; Gaps 11; Matches 25; Conservative 4; Mismatches 11; Indels 2; Gaps 11
Dd	78 SQVAAIEKKKAAAKATATAKKAIAEKKAARNAKE 119
ID	Q7W3X2 PRELIMINARY; PRT; 197 AA.
AC	Q7W3X2;
DT	01-OCT-2003 (TrEMBLrel. 25, Created)
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
CN	BPH1 OR BPP3805.
CS	Bordetella pertussis.
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Bordetella.
OX	NCBI_TaxID=519;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=12822 / ATCC BA-587;
RK	MEDLINE=22827954; PubMed=12910271;
RA	Parkhill J., Sealthina M., Preston A., Murphy L.D., Thomson N., Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L., Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A., Achtman M., Atkin R., Baker S., Basham D., Bacon N., Cherevach I., Chillingworth T., Collins M., Cronin A., Davis P., Doggett J., Fellwell T., Gobie A., Hamlin A., Hauser H., Holtzoy S., Jagels K., Leather S., Moule S., Norczek H., O'Neil S., Ormond D., Price C., Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K., Ra Sharp S., Simmonds M., Skelton V., Squares R., Squares S., Stevens K., Unwin L., Whitehead S., Barrell B.G., Maskell D.J.; "Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica.";
RL	Nat. Genet. 35:32-40(2003).
DR	EMBL; BX640435; CAE39188.1;
KM	Complete proteome.
SQ	SEQUENCE 197 AA; 19583 MW; 11D642F2BF5AE1BE CRC64;
Dy	Query Match 42.5%; Score 90.5; DB 16; Length 197; Best Local Similarity 60.5%; Pred. No. 0.13; Indels 3; Gaps 2; Matches 26; Conservative 4; Mismatches 10; Indels 3; Gaps 2
ID	Q9ZAG7 PRELIMINARY; PRT; 243 AA.
AC	Q9ZAG7;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE	Hypothetical protein lin2055.
RN	[1]

Query Match 42.3% Score 90 DB 16 Length 243;  
Best Local Similarity 45.5% Pred. No. 0.18;  
Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;

OY 1 AKKTA-KKAKKEKKKKAYKAEEAKA-----AKYEKAEEKAA 38  
Db 106 AKKAEEGAALAEKKAEEKKAEEKAADCKSGEDAAAKAAAAXKGELAEKKAEEKAA 165

OY 39 KEAYE 44  
Db 166 DKAKE 171

RESULT 13

QERL36 PRELIMINARY; PRT; 182 AA.

AC QERL36;  
DT 01-JUN-2002 (TREMBlrel. 21, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DT Histone-like protein Bph2 (Fragment).  
GN BPH2.  
OS Burkholderia multivorans.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
CX Burkholderiaceae; Burkholderia.  
OX NCBI\_Taxid=87883;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kholti A., Pleša W., Cornelis P.;  
RT "Mini-Thoprt, a new mini-transposon for in vivo protein epitope  
tagging."  
RL Submitted (FEB-2003) to the EMBL/genbank/DBJ databases.  
DR EMBL; AY089703; AAM09090.2; -  
FT NON TER 182  
SQ SEQUENCE 182 AA; 18768 MW; 243C5AECDA5A0234 CRC64;

Query Match 41.8% Score 89 DB 2 Length 182;  
Best Local Similarity 60.4% Pred. No. 0.18;  
Matches 29; Conservative 3; Mismatches 10; Indels 6; Gaps 3;

OY 1 AKKY-AKKAKEK---AKK-AVKAKEAKKAAVKAAEKAAEA 42  
Db 19 AKKTOPAKGAAAVKVVAKKTAIVKVAAKKAAPAGCAAKKVAAKKA 66

```

RESULT 14
ID Q9WMX1 PRELIMINARY; PRT; 372 AA.
AC Q9WMX1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, last annotation update)
DE Tola protein.
GN Tola.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
EX MEDLINE=96198174; PubMed=8626299;
FA Rodriguez-Herva J.U., Ramos J.,
R Rodriguez-Herva J.U., Ramos-Gonzalez M.I., Ramos J.,
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
RT 1ipoprotein (PAL) is involved in maintenance of the integrity of the
RT cell envelope."
RT J. Bacteriol. 178:1699-1706(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RA Ramos-Gonzalez J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RA Rodriguez-Herva J.U.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
EX MEDLINE=96422022; PubMed=8924639;
RA Rodriguez-Herva J.U., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RL J. Bacteriol. 178:5836-5840(1996).
DR EMBL; X74218; CAB50780.1; -.
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; F:nucleosome assembly; IEA.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR006260; TonB_C.
DR PRINTS; PRO0624; HISTONEH5.
DR TIGRFAMs; TIGR01352; tonB Cterm; 1.
SQ SEQUENCE 372 AA; 40133 MW; 67F49785EC3C0BC CRC64;

Query Match 41.5%; Score 88.5; DB 2; Length 372;
Best Local Similarity 39.7%; Pred. No. 0.39;
Matches 27; Conservative 8; Mismatches 10; Indels 23; Gaps 1.

QY 1 AKYAKAKKAKKAKYAKAAYAKAYEKAA-----AEKAA 37
DB 121 ASDAKAAEAKKAAEAKKAAEAKKAADEKKAAEKQADIAKKADEKKAABEAKKAA 180
QY 38 AKAAAYEA 45
DB 181 ABEAKKAA 188

RESULT 15
ID Q88N16 PRELIMINARY; PRT; 372 AA.
AC Q88N16;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, last annotation update)
DE Biopolymer transport protein Tola.

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 20, 2004, 23:08:29 ; Search time 29.2208 Seconds  
(without alignments)  
435.123 Million cell updates/sec

Title: US-09-816-989A-2  
Perfect score: 213  
Sequence: 1 AKYAKKAKAKAKAKAYKAA.....AKYKAAKAKAKAKAYKAA 45

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A\_Geneseq\_29Jan04.\*

1: geneseqp1980s.\*  
2: geneseqp1990s.\*  
3: geneseqp2000s.\*  
4: geneseqp2001s.\*  
5: geneseqp2002s.\*  
6: geneseqp2003as.\*  
7: geneseqp2003bs.\*  
8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	100.0	45	3	AAV82572 Copolymer
2	138	64.8	109	3	AAV82577 Copolymer
3	134.5	63.1	56	3	AAV82573 Copolymer
4	131	61.5	77	3	AAV82575 Copolymer
5	126.5	59.4	86	3	AAV82576 Copolymer
6	120.5	56.6	66	3	AAV82574 Copolymer
7	108.5	50.9	428	6	ABU27824 Protein e
8	108	50.7	35	3	AAV82571 Copolymer
9	99.5	46.7	106	2	AAV82572 Copolymer
10	98	46.0	46	2	AAV82571 Copolymer
11	98	46.0	421	6	ABU28871 High affi
12	95	44.6	323	6	ABU31397 Protein e
13	94.5	44.4	239	6	ABU31397 Protein e
14	94.5	44.4	239	6	ABU31397 Protein e
15	94	44.1	154	2	AAV82571 Copolymer
16	91	42.7	223	2	AAV82571 Copolymer
17	91	42.7	223	2	AAV82571 Copolymer
18	89.5	42.0	214	2	AAV82571 Copolymer
19	89.5	42.0	214	2	AAV82571 Copolymer
20	89.5	42.0	214	2	AAV82571 Copolymer
21	89.5	42.0	214	2	AAV82571 Copolymer
22	88.5	41.5	372	6	ABU40185 Protein e
23	87	40.8	334	4	ABG28693 Novel hum
24	86	40.4	339	2	AAV4934 Mycobacte
25	86	40.4	198	2	AAV45082 Mycobacte

26	86	40.4	198	2	AAV4936 Mycobacte
27	86	40.4	347	6	ABU38313 Protein e
28	86	40.4	347	6	ABU38313 Protein e
29	85.5	40.1	205	3	AAV82575 Copolymer
30	85	39.9	427	4	AAV82575 Copolymer
31	83	39.0	389	6	ABU39221 Protein e
32	82	38.5	165	4	AAV82571 Copolymer
33	81.5	38.3	357	6	ABV67869 Photorehab
34	81	38.0	80	5	ABV71044 Tumour ne
35	79.5	37.3	407	6	ABU47123 Protein e
36	79	37.1	212	6	ABU3811 Protein e
37	78.5	36.9	875	6	ABU2879 Protein e
38	78	36.6	157	6	ABU26064 Protein e
39	78	36.6	309	7	ABU23523 Pseudomon
40	77.5	36.4	299	4	ABV63376 Drosophi
41	77.5	36.4	376	6	ABU47848 Protein e
42	77.5	36.4	700	7	ADG95467 E. faeciu
43	77.5	36.4	2151	4	ABV60086 Drosophi
44	77	36.2	103	2	AAV34067 Histone H
45	77	36.2	103	2	AAV57365 Human his

## ALIGNMENTS

RESULT 1  
AAV82572  
ID AAV82572 standard; peptide; 45 AA.  
XX  
XX AAV82572;  
XX  
XX 28-JUL-2000 (first entry)  
XX  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.2.  
XX  
XX  
XX Copolymer; molecular weight marker; TV-marker; immune disease;  
XX Glutathione S-transferase; autoantigen; antitumor; antiproliferative;  
XX osteopontin; immunosuppressive; antitumor; antiproliferative;  
XX antidiabetic; thyromimetic; haemostatic; antiproliferative; dermatological;  
XX antidiabetic; immunosuppressive; demyelinating disease; osteoarthritis;  
XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
XX Crohn's disease; chronic immune thrombocytopenia; purpura; colitis;  
XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;  
XX Hashimoto's disease; idiopathic myxedema; myasthenia gravis;  
XX pemphigus vulgaris; systemic lupus erythematosus.  
XX  
XX Unidentified.  
XX  
XX WO200016794-A1.  
XX  
XX 06-APR-2000.  
XX  
XX 24-SEP-1999; 99WO-US022402.  
XX  
XX 25-SEP-1998; 98US-0101693P.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (TEVA-) TEVA PHARM USA INC.  
XX  
XX Gad A, Lie D;  
XX  
XX MPI; 2000-317499/27.  
XX  
XX Copolymer 1 related polypeptides used as molecular weight markers for  
XX glutathione acetate and for treatment and prevention of immune diseases.  
XX  
XX Claim 10; Page 14; 72pp; English.  
XX  
XX AAV82572 to AAV82577 represent specifically claimed copolymer molecular  
XX weight TV-marker polypeptides from the present invention. The present  
XX invention describes polypeptides (I) for determining the molecular weight  
XX of a copolymer (CP), which has an identified molecular weight and an  
XX amino acid composition corresponding to the copolymer. The polypeptides



XX Claim 10; Page 14; 72pp; English.

PS

XX

CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular

CC weight TV-marker polypeptides from the present invention. The present

CC invention describes polypeptides (I) for determining the molecular weight

CC of a copolymer (CP), which has an identified molecular weight and an

CC amino acid composition corresponding to the copolymer. The polypeptides

CC of the invention are used as molecular weight markers for glatiramer

CC acetate related tetrapolymers. The polypeptides may also be used for

CC treating and preventing immune diseases in a mammal. Autoimmune diseases

CC which may be treated include either cell-mediated or antibody-mediated

CC diseases. Such diseases include arthritic conditions, demyelinating

CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid

CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune

CC oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's

CC disease, chronic immune thrombocytopenia purpura, colitis, contact

CC sensitivity disease, diabetes mellitus, Graves disease, myasthenia gravis,

CC psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-

CC mediated diseases which can be treated include host-versus-graft disease,

CC graft-versus-host disease, and delayed-type hypersensitivity. The

CC polypeptides of the invention have defined molecular weights and physical

CC properties which are analogous to glatiramer acetate molecules, which

CC makes them ideal for use as molecular weight markers

CC

XX

SQ Sequence 56 AA;

Query Match 63.1%; Score 134.5; DB 3; Length 56;

Best Local Similarity 68.4%; Pred. No. 1.1e-07;

Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

OY 1 AKKYAKK-----AKAKA-----KKAYKAAKAKK-----AAKTEKAAKAAKAAKAAKAA 45

1 AKKYAKKAA 56

Db

RESULT 4

AAY82575

ID AAY82575 standard; peptide; 77 AA.

XX

AC AAY82575;

XX

DT 28-JUL-2000 (first entry)

XX

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.

XX

KM Copolymer; molecular weight marker; TV-marker; immune disease;

KM glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;

KM osteopathic; immunosuppressive; antithyroid; antiinflammatory;

KM antidiabetic; thrombotic; haemostatic; antipsoriatic; dermatological;

KM antinaeemic; immunosuppressive; demyelinating disease; osteoarthritis;

KM inflammatory condition; multiple sclerosis; rheumatoid arthritis;

KM Crohn's disease; chronic immune thrombocytopenia purpura; colitis;

KM diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

KM Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

KM pemphigus vulgaris; systemic lupus erythematosus.

XX

OS unidentified.

XX

PN MO200018794-A1.

XX

PD 06-APR-2000.

XX

PF 24-SEP-1999; 99WO-US022402.

XX

PR 25-SEP-1998; 98US-0101693P.

XX

PA (YEDA ) YEDA RES & DEV CO LTD.

XX

PA (TEVA-) TEVA PHARM USA INC.

XX

PI Gad A, Iis D,

XX

DR WPI; 2000-317499/27.

XX

PT Copolymer 1 related polypeptides used as molecular weight markers for

PT glatiramer acetate and for treatment and prevention of immune diseases.

XX

PS Claim 10; Page 14; 72pp; English.

XX

CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular

CC weight TV-marker polypeptides from the present invention. The present

CC invention describes polypeptides (I) for determining the molecular weight

CC of a copolymer (CP), which has an identified molecular weight and an

CC amino acid composition corresponding to the copolymer. The polypeptides

CC of the invention are used as molecular weight markers for glatiramer

CC acetate related tetrapolymers. The polypeptides may also be used for

CC treating and preventing immune diseases in a mammal. Autoimmune diseases

CC which may be treated include either cell-mediated or antibody-mediated

CC diseases. Such diseases include arthritic conditions, demyelinating

CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid

CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune

CC oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's

CC disease, chronic immune thrombocytopenia purpura, colitis, contact

CC sensitivity disease, diabetes mellitus, Graves disease, myasthenia gravis,

CC psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-

CC mediated diseases which can be treated include host-versus-graft disease,

CC graft-versus-host disease, and delayed-type hypersensitivity. The

CC polypeptides of the invention have defined molecular weights and physical

CC properties which are analogous to glatiramer acetate molecules, which

CC makes them ideal for use as molecular weight markers

CC

XX

SQ Sequence 77 AA;

Query Match 61.5%; Score 131; DB 3; Length 77;

Best Local Similarity 50.6%; Pred. No. 3.6e-07;

Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

OY 1 AKKYAKK-----AKAKA-----KKAYKAAKAKK-----AAKTEKAAKAAKAAKAAK 29

1 AKKYAKKAA 45

Db

OY 30 -KAA 45

61 YKAA 77

Db

RESULT 5

AAY82576

ID AAY82576 standard; peptide; 86 AA.

XX

AC AAY82576;

XX

DT 28-JUL-2000 (first entry)

XX

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.

XX

KM Copolymer; molecular weight marker; TV-marker; immune disease;

KM glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;

KM osteopathic; immunosuppressive; antithyroid; antiinflammatory;

KM antidiabetic; thrombotic; haemostatic; antipsoriatic; dermatological;

KM antinaeemic; immunosuppressive; demyelinating disease; osteoarthritis;

KM inflammatory condition; multiple sclerosis; rheumatoid arthritis;

KM Crohn's disease; chronic immune thrombocytopenia purpura; colitis;

KM diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

KM Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

KM pemphigus vulgaris; systemic lupus erythematosus.

XX

OS unidentified.

XX

PN MO200018794-A1.

XX

PD 06-APR-2000.

XX

PF 24-SEP-1999; 99WO-US022402.

XX

PR 25-SEP-1998; 98US-0101693P.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
PA (TEVA-) TEVA PHARM USA INC.  
XX  
XX Gad A., Lis D;  
PI  
XX WPI, 2000-317499/27.  
DR  
PT Copolymer 1 related polypeptides used as molecular weight markers for  
FT glutiramer acetate and for treatment and prevention of immune diseases.  
XX  
PS Claim 10; Page 14; 72pp; English.

CC AA82571 to AA82574 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular weight  
CC of a copolymer (CP), which has an identified molecular weight and an  
CC amino acid composition corresponding to the copolymer. The polypeptides  
CC of the invention are used as molecular weight markers for glutiramer  
CC acetate related tetrapolymers. The polypeptides may also be used for  
CC treating and preventing immune diseases in a mammal. Autoimmune diseases  
CC which may be treated include either cell-mediated or antibody-mediated  
CC diseases. Such diseases include arthritic conditions, demyelinating  
CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid  
CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune  
CC oophoritis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's  
CC disease, chronic immune thrombocytopenia purpura, colitis, contact  
CC sensitivity diseases, diabetes mellitus, Graves disease, Guillain-Barre's  
CC syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis,  
CC psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-  
CC mediated diseases which can be treated include host-versus-graft disease,  
CC graft-versus-host disease, and delayed-type hypersensitivity. The  
CC polypeptides of the invention have defined molecular weights and physical  
CC properties which are analogous to glutiramer acetate molecules, which  
CC makes them ideal for use as molecular weight markers

XX Sequence 86 AA;

SQ

Query Match 59.4%; Score 126.5; DB 3; Length 86;  
Best Local Similarity 45.3%; Pred. No.1.2e-06;  
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3

Oy 1 AKTYAKK-----AKAEKA-----KKAYGAEEKKAAKYE----- 29  
Db 1 AKTYAKKEKAYAKKAKKAKKAKKAEARAYAAEKKKAKKAEAKKAKAKKEKVEAAEA 60  
Dd 61 YKAEPAAKKAYKBAAPAAKKAAYEA 86

Oy 30 -----KAAEKAAKKAAYEA 45  
Db 61 YKAEPAAKKAYKBAAPAAKKAAYEA 86

RESULT 6  
ID AA82574  
AA82574 standard; peptide; 66 AA.

AC AA82574;  
XX  
DT 28-JUN-2000 (first entry)  
XX  
DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.  
XX  
KM Copolymer; molecular weight marker; TV-marker; immune disease;  
KM glutiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KM osteoprotic; immunosuppressive; antithyroid; antiinflammatory;  
KM antidiabetic; thymosuppressive; haemostatic; antiporiatic; dermatological;  
KM antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;  
KM inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KM Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KM diabetes mellitus; Graves disease; Guillain-Barre s syndrome; psoriasis;  
KM Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;  
KM pemphigus vulgaris; systemic lupus erythematosus.

XX	OS	Unidentified.
XX	PN	WO200018794-A1.
XX	PP	06-APR-2000.
XX	PD	
XX	PF	24-SEP-1999; 99WO-US022402.
XX	PR	25-SEP-1998; 98US-0101693P.
XX	PA	(YEDA ) YEDA RES & DEV CO LTD.
XX	PA	(TEVA-) TEVA PHARM USA INC.
XX	PI	Gad A, IIs D;
DR	WPI:	2000-317499/27.
XX	XX	Copolymer 1 related polypeptides used as molecular weight markers for
PT	glatiramer	acetate and for treatment and prevention of immune diseases.
XX	PS	Claim 10; Page 14; 72pp: English.
XX	XX	AA18251 to AA18257 represent specifically claimed copolymer molecular
CC	weight TV-marker	polypeptides from the present invention. The present
CC	invention describes	polypeptides (I) for determining the molecular weight
CC	of a copolymer (CP),	which has an identified molecular weight and an
CC	amino acid composition	corresponding to the copolymer. The polypeptides
CC	of the invention are	used as molecular weight markers for glatiramer
CC	acetate related tetrapolymers.	The polypeptides may also be used for
CC	treating and preventing	immune diseases in a mammal. Autoimmune diseases
CC	which may be treated	include either cell-mediated or antibody-mediated
CC	diseases. Such diseases	include arthritic conditions, demyelinating
CC	diseases and inflammatory	conditions, e.g. multiple sclerosis, rheumatoid
CC	arthritis, osteoarthritis,	autoimmune haemolytic anaemia, autoimmune
CC	ophoritis, autoimmune	thyroiditis, autoimmune uveoretinitis, Crohn's
CC	disease, chronic immune	thrombocytopaenia purpura, colitis, contact
CC	sensitively disease,	diabetes mellitus, Graves disease, Guillain-Barre's
CC	syndrome, Hashimoto's	disease, idiopathic myxoedema, myasthenia gravis,
CC	psoriasis, pemphigus	vulgaris, or systemic lupus erythematosus. Mediated-
CC	mediated diseases which	can be treated include host-versus-graft disease,
CC	graft-versus-host	disease, and delayed-type hypersensitivity. The
CC	polypeptides of the	invention have defined molecular weights and physical
CC	properties which are	analogous to glatiramer acetate molecules, which
CC	makes them ideal for	use as molecular weight markers
SQ	Sequence 66 AA;	
XX	Query Match	56.6%; Score 120.5; DB 3; Length 66;
XX	Best Local Similarity	71.1%; Pred. No. 4.2e-06;
XX	Matches 32; Conservative	0; Mismatches 6; Indels 7; Gaps 2;
OY	1 AKKYAKKAKKKAKKAYAAAEAKKAAEKKAAEKKAAYEA 45	
Dd	29 AKKYAKKAXKE--KKEYAANAAX-----YKFAKAAKAAKKAAYEA 66	
XX	RESULT 7	
XX	ID ABU27824	
XX	ABU27824 standard; protein; 428 AA.	
AC	ABU27824;	
XX	DT 19-JUN-2003 (first entry)	
DE	Protein encoded by Prokaryotic essential gene #13351.	
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.	
OS	Enterobacter cloacae.	
XX	MO200277183-A2.	
XX	XX	

03-OCT-2002.  
 21-MAR-2002; 2002WO-US009107.  
 21-MAR-2001; 2001US-00815242.  
 06-SEP-2001; 2001US-00948993.  
 25-OCT-2001; 2001US-0342923P.  
 08-FEB-2002; 2002US-00072851.  
 06-MAR-2002; 2002US-0362699P.  
 (EIT-) EITRA PHARM INC.  
 Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 Wall D, Trawick JD, Carr GT, Yamamoto R, Foreyth RA, Xu HH;  
 WPI; 2003-029926/02.  
 N-PSDB; ACP31694.  
 New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.  
 Claim 25; SEQ ID NO 55748; 1766bp; English.  
 The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation; (7) identifying a gene in an operon required for proliferation; (8) identifying a gene in an operon required for proliferation that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Sequence 428 AA;  
 Query Match 50.9%; Score 108.5; DB 6; Length 428;  
 Best Local Similarity 60.4%; Pred. No. 0.00058;  
 Matches 32; Conservative 2; Mismatches 8; Indels 11; Gaps 2;

1 AKKTA---KKAKKAKKAYKAAE-----AKKAKYKKAEEKAAEAA 42  
 199 AKKAADAKKAAEAAKKAQAQEAEEKAAEAAKKAEEKAAEAAEKA 251

RESULT 8  
 ID AAY82571 standard; peptide: 35 AA.  
 AC AAY82571;  
 XX

28-JUL-2000 (first entry)  
 Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.1.  
 Copolymer; molecular weight marker; TV-marker; immune disease; glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective; osteopathic; immunosuppressive; antithyroid; antiinflammatory; antidiabetic; thymostatic; haemostatic; antipsoriatic; dermatological; immunosuppressive; demyelinating disease; osteoarthritis; inflammatory condition; multiple sclerosis; rheumatoid arthritis; Crohn's disease; chronic immune thrombocytopenia purpura; colitis; diabetes mellitus; chronic immune thrombocytopenia purpura; colitis; Hashimoto's disease; idiopathic myxedema; Guillain-Barre's syndrome; psoriasis; pemphigus vulgaris; systemic lupus erythematosus.  
 Unidentified.  
 WO200018794-A1.  
 06-APR-2000.  
 24-SEP-1999; 99WO-US022402.  
 25-SEP-1999; 98US-0101693P.  
 (YEDA ) YEDA RES & DEV CO LTD.  
 (TEVA-) TEVA PHARM USA INC.  
 Gad A, Lis D;  
 WPI; 2000-317499/27.  
 Copolymer 1 related polypeptides used as molecular weight markers for glatiramer acetate and for treatment and prevention of immune diseases.  
 Claim 10; Page 14; 72pp; English.  
 AAY82571 to AAY82577 represent specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glatiramer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune coprophitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopenia purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glatiramer acetate molecules, which makes them ideal for use as molecular weight markers

Sequence 35 AA;  
 Query Match 50.7%; Score 108; DB 3; Length 35;  
 Best Local Similarity 64.4%; Pred. No. 5e-05;  
 Matches 29; Conservative 1; Mismatches 5; Indels 10; Gaps 2;

1 AKKTA---KKAKKAKKAYKAAE-----AKKAKYKKAEEKAAEAA 45  
 1 AKKTA---KKAKKAKKAYKAAE-----AKKAKYKKAEEKAAEAA 45

RESULT 9  
 ID AAR06446  
 AAR06446



ID	AA06446	standard; protein; 106 AA.
XX	AA06446;	
AC	AA06446;	
XX	25-MAR-2003	(revised)
DT	03-JAN-1991	(first entry)
XX		
DE	Recombinant copolymer 1-19, myelin basic protein analogue.	
XX		
KW	Recombinant copolymer 1; COP-1-19; myelin basic protein, MBP;	
KM	immunological activity; autoimmune encephalomyelitis; multiple sclerosis.	
OS	Synthetic.	
FN	EP983620-A.	
PD	22-AUG-1990.	
XX		
PX	16-FEB-1990;	90BP-00301700.
PR	17-FEB-1989;	89US-00312541.
PR	07-FEB-1990;	90US-00475845.
XX	(REPK ) REPLIGEN CORP.	
PA		
XX	Cook KS;	
P1		
DR	WP1: 1990-255848/34.	
DR	N-P5DB; AAO06446.	
FT	Producing genes encoding random polymers of aminoacid(s) - for producing recombinant polypeptide(s) with biological and/or immunological activity.	
XX	Disclosure; Fig 12; 25pp; English.	
PS		
CC	To improve the expression of rCOP-1 polypeptides in E. coli; genes coding for rCOP-1-19 were subcloned from pREV 2.1 to pBG3-2deltaN (deposit: 20-NV-1984 USA6931009, NMRL B-15910), a plasmid used to express Protein A. The resulting plasmids encode fusion proteins consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences. A methionine residue exists between the Protein A and rCOP-1 sequences, originating from the 5' linker sequence, in order that the COP-1 polypeptide may be cleaved from the fusion protein. rCOP-1-19 contains oligonucleotide duplexes encoding the following segments: YKK, AAF, KAK, EXA, KKX, YEK, AAA KEA, CC and AAA. The N-terminal alanine residue is left behind following CNBr cleavage of the fusion protein. The product prevents or arrests experimental autoimmune encephalomyelitis. They are used to prevent, arrest or control a demyelinating disorder, e.g. multiple sclerosis. They may also be used as additives to hair care products to confer beneficial effects on damaged hair or as supplements for diets deficient in certain amino acids. See also AAO05664. (Updated on 25-MAR-2003 to correct PA field.)	
SO	Sequence 106 AA;	
OY		
Dz		
Query Match	46.7%;	Score 99.5; DB 2; Length 106;
Best Local Similarity	58.3%;	Pred. No. 0.0013;
Matches	28; Conservative	4; Mismatches 7; Indels 9; Gaps 2
3 KYAKKA-KAEKAKAYKALAEKAKAYEKAAEKA-----AKEA 41	:	
8 KAATKYAEAEKAKAKYEAKKEAKAKAEAEAEAEKAKAKAKAEAKEA 55		
RESULT 10		
AAR28871		
ID	AAF28871	standard; peptide; 46 AA.
AC	AAR28871;	
XX		
DT	25-MAR-2003	(revised)
DT	23-MAR-1993	(first entry)
XX		

DE	High affinity macrophage mannose receptor ligand compound #5.
XX	
KW	glycopeptide; mannose; mannosylated; glycosylated; mannose receptor;
KM	macrophages; monocytcs; destroy; cytotoxicity; label; image; alter;
KW	macrophage processing of antigen; MHC restriction; inflammation;
KM	inflammatory diseases; macrophage secretory products; Crohn's disease;
KW	lygonnaires disease; mononuclear phagocytes; HIV; AIDS;
KM	lysosomal storage diseases; Gaucher's disease; asthma;
KW	alveolar macrophages metastasis; systemic macrophages; deliver;
KM	antigenic peptides; prevent transplant rejection; organ transplantation
KW	anticumour agents; cancer; toxins.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Modified-site
FT	Location/qualifiers
FT	1 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine. May also have non interfering substituents."
FT	4 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	7 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	10 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	13 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	16 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	19 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	22 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	25 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	28 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	31 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	34 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	37 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	40 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	43 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine."
FT	46 /note= "opt may have mannose, fucose, glucose or N-Ac-
FT	glucosamine. May also have non interfering substituents."
XX	
FN	WO9219248-A1.
PD	12-NOV-1992.
PF	01-MAY-1992.
PR	03-MAY-1991.
RA	(UNIV ) UNIV WASHINGTON.

XX	Stahl PD;
PI	
XX	WPI; 1992-398516/48.
DR	
XX	New high affinity mannose receptor ligand cpds. - for treating diseases
PT	mediated by macrophage activity e.g. asthma, inflammatory diseases and
PT	infectious diseases, e.g. HIV.
XX	
PS	Claim 3; Page 21; 32pp; English.
CC	
CC	This compound represents a glycopeptide effective in inhibiting the
CC	binding of labelled mannosylated BSA to mannose receptors. Mannose
CC	receptors are uniquely found on macrophages and not on monocytes.
CC	Glycopeptides such as this provide a mechanism to target macrophages
CC	specifically, to image, label, destroy or otherwise alter their antigen
CC	processing function. In addition they can be conjugated to solid supports
CC	and used to purify mannose receptors from a variety of sources. They are
CC	useful in the treatment of inflammatory diseases driven by macrophage
CC	secretory products eg. Crohn's disease; infectious diseases in which
CC	macrophages harbour replicating infectious agents eg. Legionnaires
CC	disease; viral infections involving monocuclear phagocytes eg. HIV and
CC	lysosomal storage diseases, in which macrophages are principally involved
CC	eg. Gaucher's disease; asthma mediated by alveolar macrophages; and in
CC	controlling metastasis; mediated by systemic macrophages; and in
CC	can also be used to deliver antigenic peptides as conjugates to a
CC	macrophage to marshal an immune response; also self peptides to prevent
CC	tissue transplant rejection. (Updated on 25-MAR-2003 to correct PN
CC	field.)
XX	
SQ	Sequence 46 AA;
Oy	
Dn	
Query Match	46.0%; Score 98; DB 2; Length 46;
Best Local Similarity	63.4%; Pred. No. 0.00082;
Matches 26; Conservative 1; Mismatches 14; Indels 0; Gaps 0;	
1 AKKKAKKAAEKAKKAAEKAKKAAEKAKKAAEKAKKEA 41	
2 AAKAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 42	
RESULT 11	
ABU28559	
ID ABU28559 standard; protein; 421 AA.	
XX	
AC	ABU28559;
DT	
19-JUN-2003 (first entry)	
XX	
DE	Protein encoded by Prokaryotic essential gene #14086.
XX	
KM	Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS	Escherichia coli.
XX	
PN	WO200277183-A2.
PD	
03-OCT-2002.	
XX	
PF	
21-MAR-2002; 2002WO-US0009107.	
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072651.
PR	06-MAR-2002; 2002US-0362699P.
XX	
PA	(ELIT-) ELITRA PHARM INC.
PI	
Wang L, Zamudio C, Malone C, Hasselbeck R, Ohlsen KL, Zyskind JW,	
Pi Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;	
DR WPI; 2003-029926/02.	

DR	N-PSDB; ACAA2429.
XX	New antisense nucleic acid; useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
FT	isolate candidate molecules for rational drug discovery programs.
XX	
PS	Claim 25; SEQ ID NO 56483; 1766bp; English.
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 6213 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC	polypeptide or its fragment whose expression is inhibited by the
CC	antisense nucleic acid; (4) an antibody capable of specifically binding
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC	proliferation or the activity of a gene in an operon required for
CC	proliferation; (7) identifying a compound that influences the activity of
CC	the gene product or that has an activity against a biological pathway
CC	required for proliferation, or that inhibits cellular proliferation; (8)
CC	identifying a gene required for cellular proliferation or the biological
CC	pathway in which a proliferation-required gene or its gene product lies
CC	or a gene on which the test compound that inhibits proliferation of an
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC	compound's activity; (11) a culture comprising strains in which the gene
CC	product is overexpressed or underexpressed; (12) determining the extent
CC	to which each of the strains is present in a culture or collection of
CC	strains; or (13) identifying the target of a compound that inhibits the
CC	proliferation of an organism. The antisense nucleic acids are useful for
CC	identifying proteins or screening for homologous nucleic acids required
CC	for cellular proliferation to isolate candidate molecules for rational
CC	drug discovery programs, or for screening homologous nucleic acids
CC	required for proliferation in cells other than S. aureus, S. typhimurium,
CC	K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC	the largest prokaryotic essential genes. Note: The sequence data for this
CC	patent did not form part of the printed specification, but was obtained
CC	in electronic format directly from WIPO at
CC	fhp.wipo.int/pub/published_pct_sequences
SQ	Sequence 421 AA;
Query Match	46.0%; Score 98; DB 6; Length 421;
Best Local Similarity	65.3%; Pred.No.0.008; Mismatch 9; Indels 6; Gaps 3
Matches	32; Conservative 2; Mismatches 9; Indels 6; Gaps 3
Dy	2 KKTAKKAAEKKA---KK--AVYAAEBAKKAAYTERAEEAAAKEAAVEEA 45   Db      KKAAEGQAEEKAADKKAAAEKKAADKKAAA-AAVAAEKAAAAKAAAEA 294
RESULT 12	
ID ABU31397 standard; protein; 323 AA.	
ABU31397	
AC ABU31397;	
DT 19-JUN-2003 (first entry)	
DE Protein encoded by Prokaryotic essential gene #16924.	
DD Antisense; prokaryotic essential gene; cell proliferation; drug design.	
FM Klebsiella pneumoniae.	
OS MO200277183-A2.	
FN WO200277183-A2.	
PD 03-OCT-2002.	
PF 21-MAR-2002; 2002WO-US009107.	
PR 21-MAR-2001; 2001US-00815242. 06-SEP-2001; 2001US-00948993.	



XX MO20027183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WC-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362659P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zyskind JW,  
XX Wall D, Trawick JD, Carr GT, Yamamoto R, Foreyth RA, Xu HH,  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA36489.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 60543; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 613 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is specifically binding  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 239 AA;  
XX Query Match 44.4%; Score 94.5; DB 6; Length 239;  
XX Best Local Similarity 56.2%; Pred. No. 0.011;  
XX Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;  
QY 2 KKTAKKAKKAKK--AYAAAKK-AAKYEKAAKAAKAAAYE 44  
DB 124 KAAAEKAEADKKQEDAVKAAAKKQEAEEKAAADKAAAEKAAAE 171  
RESULT 15  
AAR06445  
ID AAR06445 standard; protein; 154 AA.

XX AAR06445;  
XX 25-MAR-2003 (revised)  
XX 03-JAN-1991 (first entry)  
XX Recombinant copolymer 1-77; myelin basic protein analogue.  
XX Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;  
XX immunological activity; autoimmune encephalomyelitis; multiple sclerosis.  
XX Synthetic.  
XX EP83620-A.  
XX 22-AUG-1990.  
XX 16-FEB-1990; 90EP-00301700.  
XX 17-FEB-1989; 89US-00312541.  
XX 07-FEB-1990; 90US-00473845.  
XX (REPK ) REPLIGEN CORP.  
XX Cook KS;  
XX WPI; 1990-255848/34.  
XX N-PSDB; AAQ05664.  
XX Producing genes encoding random polymers of aminoacid(s) - for producing  
XX PT recombinant polypeptide(s) with biological and/or immunological activity.  
XX Disclosure; Fig 11; 25pp; English.  
XX To improve the expression of rCOP-1 polypeptides in *E. coli*, genes coding  
XX CC for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2deltaN (deposit: 20-  
XX CC NOV-1984 US4691009, NRL B-15910), a plasmid used to express Protein A.  
XX CC The resulting plasmids encode fusion proteins consisting of beta-  
XX CC glucuronidase, Protein A, and rCOP-1 sequences. A methionine residue  
XX CC occurs between the Protein A and rCOP-1 sequences, originating from the  
XX CC 5' linker sequence, in order that the COP-1 polypeptide may be cleaved  
XX CC from the fusion protein. rCOP-1-77 contains oligonucleotide duplexes  
XX CC encoding the following segments: YKK, EAE, KAK, AAK, and AAA. The N-  
XX CC terminal alanine residue is left behind following CNBr cleavage of the  
XX CC fusion protein. The product prevents or arrests experimental autoimmune  
XX CC encephalomyelitis. They are used to prevent, arrest or control a  
XX CC demyelinating disorder, e.g. multiple sclerosis. They may also be used as  
XX CC additives to hair care products to confer beneficial effects on damaged  
XX CC hair or as supplements for diets deficient in certain amino acids. See  
XX CC also AAQ05665. (Updated on 25-MAR-2003 to correct PA field.)  
XX Sequence 154 AA;  
XX Query Match 44.1%; Score 94; DB 2; Length 154;  
XX Best Local Similarity 60.5%; Pred. No. 0.0077;  
XX Matches 26; Conservative 2; Mismatches 9; Indels 6; Gaps 2;  
QY 2 KKTAKKAKKAKK--AYAAAKK-AAKYEKAAKAAKAAAYE 44  
DB 105 KKYKAKKAKKAKK--KAEAEKA---KAAAEKAKKAEAEYK 141  
Search completed: April 20, 2004, 23:15:46  
Job time : 31.2208 secs





US-09-816-989A-4  
Sequence 4, Application US/09816989A  
Patent No.: US20020115103A1  
GENERAL INFORMATION:  
APPLICANT: Gad, Alexander  
APPLICANT: Lie, Doris  
TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARKERS  
FILE REFERENCE: AND FOR THERAPEUTIC USE  
CURRENT APPLICATION NUMBER: US/09/816, 989A  
CURRENT FILING DATE: 2001-03-23  
PRIORITY APPLICATION NUMBER: 60/101,693  
PRIOR FILING DATE: 1998-09-25  
PRIOR APPLICATION NUMBER: PCT/US99/22402  
PRIOR FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 66  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURES:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide

US-09-816-989A-4

Query March 56.6%; Score 120.5; DB 9; Length 66;  
Best Local Similarity 71.1%; Pred. No. 3.1e-06;  
Matches 32; Conservative 0; Mismatches 6; Indels 7; Gaps 2;

Qy 1 AKKYAKKAEKKAKAYKAABEKATYERKAAAEKAAEAYEA 45  
Db 29 AKKYAKKAAAE--KXEYAAAARX-----YRAEAKKAAAEAYEA 66

RESULT 7  
US-10-282-122A-55748  
Sequence 55748, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA 034A  
CURRENT APPLICATION NUMBER: US/10/282, 122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 55748
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-55748

Query Match      50.9%; Score 108.5; DB 12; Length 428;
Best Local Similarity   60.4%; Pred.No.0.00047;
Matches          32; Conservative    2; Mismatches     8; Indels    11; Gaps       2;

Oy              1 AKKVA-----KKAAEKAKKAYVPAE-----AKKAATYEKAAREKAAKEAA 42
                ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db             199 AKKAADQKRAEALEAKKTAQAEKKAALAAEAARKAAAAAKEAAAEEKAAAEKAA 251

RESULT 8
US-09-816-989A-1
; Sequence 1, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lie, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MA
; TITLE OF INVENTION: AND FOR THERAPEUTIC USE
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-1

Query Match      50.7%; Score 108; DB 9; Length 35;
Best Local Similarity   64.4%; Pred.No.3.6e-05;
Matches          29; Conservative    1; Mismatches     5; Indels    10; Gaps        2;

Oy              1 AKKYAKKAKAEKAKKAYRPAAEKKAKAYEKAAEFKRAAAAEAAVEA 45
                ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db             1 AKKXAKKEKA--AKKAYK-----KKAKAPAAAEAKAAVEA 35

RESULT 9
US-10-282-122A-56483
; Sequence 56483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haaslebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trewick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```

, RESULT 10
, US-10-282-122A-59321
, Sequence 59321, Application US/10282122A
, Publication NO. US20040028129A1
, GENERAL INFORMATION:
, APPLICANT: Wang, Liangsu
, APPLICANT: Zamudio, Carlos
, APPLICANT: Malone, Cheryl
, APPLICANT: Haselbeck, Robert
, APPLICANT: Ohlsen, Kari
, APPLICANT: Zyakind, Judith
, APPLICANT: Wall, Daniel
, APPLICANT: Trawick, John
, APPLICANT: Carr, Grant
, APPLICANT: Yamamoto, Robert
, APPLICANT: Forsyth, R.
, APPLICANT: Xu, H.
, TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
, FILE REFERENCE: EUTPR, 034A
, CURRENT APPLICATION NUMBER: US/10/282,122A
, CURRENT FILING DATE: 2003-02-20
, PRIOR APPLICATION NUMBER: 60/191,078
, PRIOR FILING DATE: 2000-03-21
, PRIOR APPLICATION NUMBER: 60/206,848
, PRIOR FILING DATE: 2000-05-23
, PRIOR APPLICATION NUMBER: 60/207,727
, PRIOR FILING DATE: 2000-05-26
, PRIOR APPLICATION NUMBER: 60/230,335
, PRIOR FILING DATE: 2000-09-06
, PRIOR APPLICATION NUMBER: 60/230,347
, PRIOR FILING DATE: 2000-09-09
, PRIOR APPLICATION NUMBER: 60/242,578
,

```

```

RESULT 11
US-10-282-122A-60543
Sequence 60543, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Mall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITPR 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/250,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Reminding Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 60543
LENGTH: 239

```





APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 64817  
LENGTH: 214  
TYPE: PRF  
ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64817

Query Match 42.0%; Score 89.5; DB 12; Length 214;  
Best Local Similarity 55.6%; Pred. No. 0.026;  
Matches 25; Conservative 4; Mismatches 13; Indels 3; Gaps 1;

QY 1 AKVAKKAKAERAKKAYKAAEKAKAERAKKAAEKAAEAYEA 45  
DB 111 AKVAKKAKAERAKKAYKAAEKAKAERAKKAAEKAAEAYEA 152

Search completed: April 20, 2004, 23:27:17  
Job time : 23.2078 secs







QY 1 AKKAAKAEKAY---AKKAAKAEKAYAKKAAKAYAAE-----AKKAAKAEKAY 49  
 DB 1067 AKKLEAAEKKECEERAKKEAAEKKECEERAKKLEAAEKKECEERAKKEAAEKKE 1126  
 QY 50 AKKAAKAEKAYAKKAAKAYAKKAAKAEKKEEYAAAEKKEEAAKAAKAAKAAKAAEAA 106  
 DB 1127 CEAAKKEKAA-AREKKEAAKKEKKE-ATEKOKCAAAKKEKKEAAKKEKKECAAA 1179

## RESULT 11

HSTUR1P  
 histone H1, gonadal - sea urchin (Parechinus angulosus)  
 C/Species: Parechinus angulosus (angulate urchin)  
 C/Date: 31-Mar-1980 #sequence\_revision 31-Mar-1980 #text\_change 16-Feb-1997  
 C/Accession: A91090; A91091; A02586  
 R/Strickland, W.N.; Strickland, M.; de Groot, P.C.; von Holt, C.; Wittmann-Liebold, B.  
 Eur. J. Biochem. 194, 559-566, 1980  
 A/Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus angulosus  
 A/Reference number: A91090; PMID:80156831; PMID:6767609  
 A/Contents: sequence of residues 1-84  
 A/Accession: A91090  
 A/Molecule type: protein  
 A/Residues: 1-248 <STR>  
 R/Strickland, W.N.; Strickland, M.; Brandt, W.F.; von Holt, C.; Lehmann, A.; Wittmann-Liebold, B.  
 Eur. J. Biochem. 194, 567-578, 1980  
 A/Title: The primary structure of histone H1 from sperm of the sea urchin Parechinus angulosus  
 A/Reference number: A91091; PMID:80156832; PMID:7363905  
 A/Accession: A91091  
 A/Molecule type: protein  
 A/Residues: 80-248 <STR>  
 A/Note: 144-Arg was also found  
 A/Suprafamily: histone H1  
 C/Keywords: DNA binding; nucleosome; sperm

Query Match 32.9%; Score 170.5; DB 1; Length 248;  
 Best Local Similarity 46.0%; Pred. No. 0.00017;  
 Matches 52; Conservative 9; Mismatches 41; Indels 11; Gaps 2;

QY 2 KKYAKKAEKAYAKKAAKAEKAYAKKAAKAAEAKKAAKAAKAAKAAKAAKAAKAAK 61  
 DB 120 KKAATTSAAAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 179  
 QY 62 KAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 103  
 DB 180 KAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 232

## RESULT 12

T09127  
 probable erythrocyte-binding protein MAEBL - Plasmodium yoelii  
 C/Species: Plasmodium yoelii  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
 C/Accession: T09127  
 R/Kapke, S.H.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.  
 Proc. Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998  
 A/Title: A family of chimeric erythrocyte binding proteins of malaria parasites.  
 A/Reference number: Z16577; PMID:98115903; PMID:9448314  
 A/Accession: T09127  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-1701 <KAP>  
 A/Cross-references: EMBL:AF031886; NID:g2947227; PID:g2947228  
 A/Experimental source: subspecies yoelii; strain YM  
 C/Genetics:  
 A/Gene: maeb1  
 A/Introns: 62/1; 1648/1; 1674/2; 1697/1  
 C/Keywords: alternative splicing; cell binding; erythrocyte invasion

Query Match 32.9%; Score 170.5; DB 2; Length 1701;  
 Best Local Similarity 53.5%; Pred. No. 0.00068;  
 Matches 54; Conservative 11; Mismatches 31; Indels 5; Gaps 5;

QY 5 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 63  
 DB 1205 AKKAAEE-EKKAAVAV-KAEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 1261  
 QY 64 EAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 104  
 DB 1262 EKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 1301

## RESULT 13

A28456  
 histone H1.10 - chicken  
 C/Species: Gallus gallus (chicken)  
 C/Date: 28-Aug-1989 #sequence\_revision 28-Aug-1989 #text\_change 04-Mar-2000  
 C/Accession: A28456  
 R/Coles, L.S.; Robins, A.J.; Madley, L.K.; Wells, J.R.B.  
 J. Biol. Chem. 262, 9656-9663, 1987  
 A/Title: Characterization of the chicken histone H1 gene complement. Generation of a cDNA library  
 A/Reference number: A28455; PMID:87250632; PMID:3597432  
 A/Accession: A28456  
 A/Molecule type: DNA  
 A/Residues: 1-220 <COL>  
 A/Cross-references: GB:M17018; NID:G211834; PID:AAA48788.1; PID:G211835  
 C/Suprafamily: histone H1  
 C/Keywords: acetylated amino end; chromosomal protein; DNA binding; nucleosome; nucleus  
 F/2/Modified site: acetylated amino end (Ser) (in mature form) #status predicted

Query Match 31.8%; Score 165; DB 2; Length 220;  
 Best Local Similarity 46.4%; Pred. No. 0.00034;  
 Matches 52; Conservative 11; Mismatches 41; Indels 8; Gaps 3;

QY 1 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 53  
 DB 109 SKKPGVKKAPKPKTPAKKPKKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 168  
 QY 54 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 104  
 DB 169 AKSPKATATAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 220

## RESULT 14

S59589  
 histone H1 - Chlamydomonas reinhardtii  
 C/Species: Chlamydomonas reinhardtii  
 C/Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 23-Jul-1999  
 C/Accession: S59589; S62122  
 R/Fabry, S.; Mueller, K.; Lindauer, A.; Park, P.B.; Cornelius, T.; Schmitt, R.  
 Curr. Genet. 28, 333-345, 1995  
 A/Title: The organization structure and regulatory elements of Chlamydomonas histone ger  
 A/Reference number: S59581; PMID:96120862; PMID:8590479  
 A/Accession: S59589  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-231 <PAB>  
 A/Cross-references: EMBL:U16726  
 A/Note: the authors did not translate the codon for residue 1  
 R/Fabry, S.; Ford, C.; Lindauer, A.; Mueller, K.; Cornelius, T.; Schmitt, R.  
 submitted to the EMBL Data Library, October 1994  
 A/Description: The organization, structure and controlling elements of Chlamydomonas histone H1  
 A/Reference number: S62122  
 A/Accession: S62122  
 A/Molecule type: DNA  
 A/Residues: 1-173, 'P', 174-231 <PAB>  
 A/Cross-references: EMBL:U16726; NID:g571479; PID:AAA8452.1; PID:g571480  
 C/Genetics:  
 A/Introns: 62/3; 101/3  
 A/Suprafamily: histone H1  
 C/Keywords: chromosomal protein; DNA binding; nucleosome

Query Match 31.7%; Score 164.5; DB 2; Length 231;  
 Best Local Similarity 45.9%; Pred. No. 0.00038;  
 Matches 56; Conservative 10; Mismatches 39; Indels 17; Gaps 5;

Qy	2	KXTAKA	---	EKVYAKKAAAKAEKKYAKKGA	----	KYKAAEAKKAKAAKXYAAEA	54
Db	100	KKAATKAATPKAAAPKEGAVKTKAPKGESEKPKSAKAEKPKTEGEKTAAPAK	159				
Qy	55	KAKKEAVYAEKKYAAFAA	----	KAEKKEVAAE	----	AKKAFAA	104
Db	160	AEEKPKAAKPAKATTTTAAAAKPKAEKPKPAAAKPAAEKKPKAAKPKAEKKAAKPAAKK	219				
Qy	105	AA	106				
Db	220	SA	221				

## RESULT 15

cgcr-4 protein Chlamydomonas reinhardtii (fragment)  
C/Species: Chlamydomonas reinhardtii  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 21-Jul-2000  
C/Accession: S19113; S14466  
R/Wakarchuk, W.W.; Mueller, F.W.; Beck, C.F.  
Plant Mol. Biol. 18, 143-146, 1992  
A/Title: Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex arrangements  
A/Reference number: S19113; MUID:92119224; PMID:1731966  
A/Accession: S19113  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-265 <WAK>  
A/Cross-references: EMBL:X17208; NID:g18136; PIDN:CAA35080.1; PID:g18137  
C/Genetics:  
A/Gene: cgcr-4

Query Match	31.6%	Score 164	DB 2	Length 265
Best Local Similarity	39.4%	Pred. No. 0.00045		
Matches 50	Conservative 2	Mismatches 38	Indels 18	Gaps 2

[illegible]

Search completed: April 20, 2004, 23:18:23  
Job time : 19.1104 secs







RX MEDLINE=97113525; PubMed=8955385;  
 RA Dennis U.J., Lafontaine E.R., Sokol P.A.;  
 RT "Identification and characterization of the tolpa genes of  
 RT Pseudomonas aeruginosa.";  
 RL J. Bacteriol. 178:7059-7068 (1996).  
 RN [2]  
 RP REVISIONS TO N-TERMINUS.  
 RA Duan K., Sokol P.A.;  
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stever C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964 (2000).  
 CC -1 FUNCTION: Involved in the tonB-independent uptake of proteins (By  
 CC similarity).  
 CC -1 SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane  
 CC (potential).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U39558; AAC44660.2; -;  
 CC EMBL: AE004530; AAG04360.1; -;  
 CC PIR: E83525; E83525.  
 CC InterPro: IPR006260; TonB\_C.  
 CC TIGRFAMs: TIGR01352; tonB\_C term; 1.  
 CC Transport, Protein transport; Transmembrane; Repeat; Inner membrane;  
 CC Complete proteome.  
 CC DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).  
 CC TRANSSEM 17 37 POTENTIAL.  
 CC DOMAIN 38 347 PERIPLASMIC (POTENTIAL).  
 CC DOMAIN 209 216 POLY-ALA.  
 CC SEQUENCE 347 AA; 37935 MW; EEDDA804AA095945 CRC64;  
 SQ  
 Query Match 34.5%; Score 179; DB 1; Length 347;  
 Best Local Similarity 49.1%; Pred. No. 1.9e-05;  
 Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;  
 Oy 6 KKAEEAKYAKKAAKKEK-----KAYAKKEAKYAKKAAKKAAYAKKAAKKEAKK 59  
 ID MST1\_DROHY STANDARD; PRT; 344 AA.  
 AC Q08695;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Axoneme-associated protein ms101(1).  
 GN MS101(1).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Db 159 A-EDEAKK--KAEDAKKK-AAEDAKKKAAAEAKKAAAEAKKAAVEA 204  
 Oy 60 AYKAAKKYAAKAEKKEVYAAAEKKAEEAKKAAKAYKAAKAAEAAYEA 109  
 ID MST101(1).  
 AC Q08695;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Axoneme-associated protein ms101(1).  
 GN MS101(1).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Db 159 A-EDEAKK--KAEDAKKK-AAEDAKKKAAAEAKKAAAEAKKAAVEA 204

OC Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_Taxid=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RX MEDLINE=94200512; PubMed=8150205;  
 RA Neesen J., Buemann H., Heinlein U.A.O.;  
 RT "The Drosophila hydei gene Dms101(1) encodes a testis-specific,  
 RT repetitive, axoneme-associated protein with differential abundance in  
 RT X chromosomal deletion mutant flies.";  
 RL Dev. Biol. 162:414-425 (1994).  
 CC -1 FUNCTION: Possible structural role in the sperm tail. It is  
 CC associated with axonemal structures.  
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
 CC -1 TISSUE SPECIFICITY: Testis. Located in spermatocytes and  
 CC spermatid bundles.  
 CC -1 DOMAIN: The predominant structure is alpha-helical.  
 CC -1 POLYMORPHISM: The number of repeats varies between strains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X73480; CAA51875.1; -;  
 CC PIR: S34153; S34153.  
 CC FlyBase: Fgn0011816; Dhyd\ms101(1).  
 CC GO: GO:0005737; Cytoplasm; IDA.  
 CC DR GO: GO:0005198; Filamentous molecular activity; IEP.  
 CC DR GO: GO:0007288; Paxoneme assembly; IEP.  
 CC KW Sperm; Repeat; Multigene family.  
 CC FT DOMAIN 58 337  
 CC K-K-K-C-X-E-X-A-[KQ]-K-X-X-E-X-A-X.  
 SQ SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;  
 Query Match 33.8%; Score 175.5; DB 1; Length 344;  
 Best Local Similarity 49.2%; Pred. No. 3.1e-05;  
 Matches 58; Conservative 8; Mismatches 35; Indels 17; Gaps 5;  
 Oy 1 AKKAAKKEK--AYAKKAAKKEKAYAKKAAKAYKAAE-----AKKAAKAAK 49  
 ID MST2\_DROHY STANDARD; PRT; 1391 AA.  
 AC Q08695;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Axoneme-associated protein ms101(2).  
 GN MS101(2).  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_Taxid=7224;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=95045538; PubMed=7957199;  
 RA Neesen J., Padmanabhan S., Buemann H.;  
 RT "Randomly arranged repeats of a novel highly charged 16-amino-acid  
 RT motif representing the major component of the sperm-tail-specific  
 RT axoneme-associated protein family Dms101 form extended  
 RT alpha-helical rods within the extremely elongated spermatzoa of



DR InterPro: IPR003216; Linkerhist\_N.  
 DR Pfam: PF00538; linker\_histone; I.  
 DR PRINTS: PR00624; HISTONE5.  
 DR ProDom: PD00373; Linkerhist\_N; 1.  
 DR SMART: SM00526; H15; 1.  
 DR Chromosomal protein; Nuclear protein; Multigene family.  
 SO SEQUENCE 211 AA; 22169 MW; 9F214581334BBE7A CRC64;

Query Match 32.7%; Score 169.5; DB 1; Length 211;  
 Best Local Similarity 48.7%; Pred. No. 5.1e-05;  
 Matches 55; Conservative 8; Mismatches 41; Indels 9; Gaps 4;

QY 3 KYAKAE-KAYAKKA-KAAKEKAYAKKAAKAAK-KKAAEKYAKKAAK-- 55  
 DB 8 KLGKKEKSDQKAPDAKAKAKLAKKEKKEKKAASKAKKAKSKTTTKV 147  
 QY 56 --AKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKK 106  
 DB 148 KPAKAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKKAPAKK 200

## RESULT 8

MS13\_DROHY STANDARD; PRT; 275 AA.

AC 001395;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Anoxemia-associated protein met101(3).  
 GN MS101(3) OR DMST101.  
 OS Drosophila hydei (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7224;

SEQUENCE FROM N.A.

RA MEDLINE=99372678; PubMed=10445506;  
 RA Niesen J., Heinlein U.A.O., Heinz Glatzer K., Buemann H.;  
 RA "Proteins have been arranged repeats of a highly charged 16-amino-  
 RT acid motif encoded by the Dmst101 gene family are structural  
 RT components of the outer sheath of the extremely elongated sperm tails  
 RT of Drosophila hydei.";  
 RL Dev. Growth Differ. 41:93-99(1999).  
 CC -1- FUNCTION: Possible structural role in the sperm tail (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: Testis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/submit/> or  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL: U85627; AAB51369.1;  
 DR FLYBase: FBgn0020732; Dmst101(3).  
 DR GO: GO:0005737; Cytoplasm; ISS.  
 DR GO: GO:0005198; F: structural molecule activity; ISS.  
 DR GO: GO:0007288; P: axoneme assembly; ISS.  
 KM Sperm; Repeat; Multigene family.  
 FT DOMAIN 64 255  
 FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF  
 FT X-(KQ)-K-C-(AE)-E-X-A-[X]-K-X-X-X-X-X-  
 FT [AE]-X.

SEQUENCE 275 AA; 30436 MW; 76BAAYB2A2DFJ32C CRC64;

Query Match 32.7%; Score 169.5; DB 1; Length 275;  
 Best Local Similarity 48.2%; Pred. No. 6.4e-05;  
 Matches 54; Conservative 16; Mismatches 29; Indels 13; Gaps 6;

QY 1 AKKAYAKKAYAKK-AKKAYAKKAYAKKAYAKKAYAKKAYAKK- 57

DB 51 AEDVKKCEBAANKKCAKAAKKEKAAK--KCAKAAKKEKAAKCAKAAKE 106  
 QY 58 KENYAKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKKAYAKK 108  
 DB 107 OEA--AKKKCAKAAKKEK--AAKCCAKAAKKEKAAKKEKCAAKK 153

## RESULT 9

FAU DROME STANDARD; PRT; 668 AA.

AC Q9VGX3; Q9SG18; Q9VGX1; Q9VGX2; Q9Y0F9;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Anoxia upregulated protein.  
 GN FAU OR G6544.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7227;

SEQUENCE FROM N.A.

RA STRAIN=Canton-S; TISSUE=Head;  
 RA MEDLINE=99097004; PubMed=9878744;  
 RA Ma E., Xu T., Haddad G.G.;  
 RA "Gene regulation by O2 deprivation: an anoxia-regulated novel gene in  
 RT Drosophila melanogaster.";  
 RL Brain Res. Mol. Brain Res. 63:217-224(1999).  
 [2]

SEQUENCE FROM N.A.

RA STRAIN=Beckley;  
 RA MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Chang M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Achyutani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu B., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bonkayov S.,  
 RA Borokova D., Botchan M.R., Bouck U., Brockstein P., Brotler P.,  
 RA Burris K.C., Busan D.A., Butler H., Davenport L.B., Davies P.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,  
 RA Dodson K.U., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.U., Evangelista C.C., Ferraz C., Fertala S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalili M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.E., McPherson D.,  
 RA Merklow S.G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rhine R.C., Remington K., Saunders R.D.C., Scheel M.P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stappleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye T., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 [3]







CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M17019; AAA48789.1; -  
 DR PIR; B28456; B28456.  
 DR PDB; 1GHC; 31-AUG-94.  
 DR InterPro; IPR005818; Histone\_H1/H5.  
 DR InterPro; IPR005819; Histone\_H5.  
 DR Pfam; PF00538; linker histone; 1.  
 DR PRINTS; PR00624; HISTONEH5.  
 DR SMART; SM00526; H15; 1.  
 DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
 KW Acetylation; 3D-structure.  
 FT INIT\_MET 0  
 FT MOD\_RES 1 13 ACETYLATION (BY SIMILARITY).  
 FT DOMAIN 40 113 GLOBULAR.  
 FT HELIX 44 54  
 FT TURN 62 63  
 FT HELIX 66 68  
 FT HELIX 81 87  
 FT TURN 88 90  
 FT HELIX 91 94  
 FT TURN 95 95  
 FT TURN 104 105  
 FT TURN 104 105  
 SQ SEQUENCE 224 AA; 22397 MW; D3D057CB97865CAF CRC64;

Query Match 31.4%; Score 163; DB 1; Length 224;  
 Best Local Similarity 43.8%; Pred. No. 0.0014;  
 Matches 49; Conservative 13; Mismatches 42; Indels 8; Gaps 2;

QY 1 AKKYAKKAEKAYAKAKAEK-----AYAKKEAYAKAEKKAKAEKAYAKAEK 53  
 DB 112 SKKSEVEKAKPKKAKSAKPKKPAKPAKPAKPAKPAKPAKPAKPAKPAK 171  
 54 AYAKKEAYAKAEKAYAKAKAEKAYAA-AYAKKAEKAYAKAEKAYAKAEK 104  
 DB 172 AKSPKVKTKAVKPKKAVAKSPAKKAVKPKKPAKPAKPAKPAKPAK 223

RESULT 13  
 ID H1\_ONCMY STANDARD; PRT; 206 AA.  
 AC P06350; P83374; (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Histone H1 (Contains; Oncorhynchus II).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Procarangopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=85264847; PubMed=6443128;  
 RA Mezquita U., Connor W., Winkfein R.J., Dixon G.H.;  
 RA "An H1 histone gene from rainbow trout (Salmo gairdneri).";  
 RT J. Mol. Evol. 21:209-219 (1985).  
 RL [2]  
 RP SEQUENCE OF 138-154, FUNCTION OF ONCORHYNCHIN II, AND MASS  
 RP SPECTROMETRY OF 138-206.  
 RC TISSUE=Skin mucos;  
 RX MEDLINE=22850086; PubMed=12969798;  
 RX Fernandez J.M.O., Molle G., Kemp G.D., Smith V.J.;  
 RA "Isolation and characterisation of oncorhynchin II, a histone H1-  
 RT derived antimicrobial peptide from skin secretions of rainbow trout,  
 RN Oncorhynchus mykiss";  
 RL Dev. Comp. Immunol. 26:127-138(2004).

CC -1- FUNCTION: Histones H1 are necessary for the condensation of  
 CC nucleosome chains into higher order structures.  
 CC -1- FUNCTION: Oncorhynchin II has antibacterial activity against Gram-  
 CC positive and Gram-negative bacteria at submicromolar  
 CC concentrations. Potentially important role in mucosal defense.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Oncorhynchin II is secreted.  
 CC -1- TISSUE SPECIFICITY: Oncorhynchin II is expressed in skin.  
 CC -1- MASS SPECTROMETRY: MW=7195.3; METHOD=MALDI; RANGE=138-206.  
 CC -1- MISCELLANEOUS: Is thermostable up to 80 degrees Celsius.  
 CC -1- SIMILARITY: Belongs to the histone H1/H5 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X02624; CAB37646.1; -  
 DR PIR; A02584; HSTR1R.  
 DR HSSP; P08287; 1GHC.  
 DR InterPro; IPR005818; Histone\_H1/H5.  
 DR InterPro; IPR005819; Histone\_H5.  
 DR Pfam; PF00538; linker histone; 1.  
 DR PRINTS; PR00624; HISTONEH5.  
 DR SMART; SM00526; H15; 1.  
 DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
 KW Acetylation; Antibiotic.  
 FT INIT\_MET 0  
 FT CHAIN 1 206 HISTONE H1.  
 FT CHAIN 138 206 ONCORHYNCHIN II.  
 FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
 FT DOMAIN 27 100 GLOBULAR.  
 FT SITE 137 138 CLEAVAGE (PROBABLE).  
 SQ SEQUENCE 206 AA; 20672 MW; 72C440798066716C CRC64;

Query Match 30.9%; Score 160.5; DB 1; Length 206;  
 Best Local Similarity 47.7%; Pred. No. 0.00019;  
 Matches 52; Conservative 9; Mismatches 35; Indels 13; Gaps 4;

QY 1 AKKYAK-----KAEKAYAKAKAEKAYAKAEKAYAKAEKAYAKAEK 55  
 DB 105 AKKPAKKAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 164  
 56 AKKEAYKAEKAYAKAKAEKAYAAAEKAEKAEKAYAKAEKAYAKAEK 104  
 DB 165 VKKPA--AAAK--KAASPK--ATKAKPKPAKPAKPAKPAKPAK 205

RESULT 14  
 ID H15\_HUMAN STANDARD; PRT; 225 AA.  
 AC P16401; Q14529;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Histone H1.5 (Histone H1a).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE  
 RC TISSUE=Skin;  
 RX MEDLINE=9010391; PubMed=2613692;  
 RA Ohe Y., Hayashi H., Iwai K.;  
 RA "Human spleen histone H1. Isolation and amino acid sequences of three  
 RT minor variants, H1a, H1c, and H1d.";  
 RL J. Biochem. 106:844-857(1989).  
 RL [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=97183654; PubMed=9031620;  
 RA Albig W., Meergans T., Doenecke D.;  
 RT "Characterization of the H1.5 gene completes the set of human H1  
 RT subType genes."; Gene 184:141-148(1997).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22269985; PubMed=12408966;  
 RA Marzluft W.F., Congidi P., Woods K.R., Jin J., Maltais L.C.;  
 RT "The human and mouse replication-dependent histone genes.";  
 RL Genomics 80:487-498(2002).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RA Wild A.;  
 RL Submitted (DRC-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Histones H1 are necessary for the condensation of  
 CC nucleosome chains into higher order structures.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the histone H1/H5 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, X83509; CA58498.1; -;  
 DR EMBL, AF531304; AA006704.1; -;  
 DR EMBL, Z98744; CAB1421.1; -;  
 DR PIR, S5160; S5160.  
 DR HSSP, P08287; IGHC.  
 DR Genew; HGNC:4719; HIST1H1B.  
 DR MIM, 142711; -;  
 DR GO, GO:0000786; C:nucleosome; NAS.  
 DR GO, GO:0003677; F:DNA binding; NAS.  
 DR GO, GO:0007001; P:chromosome organization and biogenesis (sen. .); NAS.  
 DR GO, GO:0006334; P:nucleosome assembly; NAS.  
 DR InterPro: IPR005818; Histone H1/H5.  
 DR InterPro: IPR005819; Histone H5.  
 DR Pfam: PF00538; linker histone 1.  
 DR PRINTS; PR00624; HISTONEH5.  
 DR SMART; SW00526; HIS; 1.  
 KM Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;  
 KM Acetylation.  
 FT INIT MET 0 0  
 FT MOD RES 1 1  
 FT CONFLICT 215 217 MISSING (IN REF. 1).  
 FT SEQUENCE 225 AA; 26CD4ALESD463CDA CRC64;  
 SQ  
 Query Match 30.9%; Score 160.5; DB 1; Length 225;  
 Best Local Similarity 46.8%; Pred. No. 0.00021;  
 Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;  
 QY 1 AKTKAKAEKAYAKKAAKAEKAYAKKAEKAYAKA-----EKKAKAEKAYAKAEAK 55  
 DB 118 AKPKAKAGAKAKKPKAGATPKK--AKKAAGAKKAVKTPKAKKPAAGVKKVAKSPKK 175  
 QY 56 AKKEAYAAEKYAKKAAKAEKKEYYAAAEAKKAEAKKAYAKKAAKAAKAAK 104  
 DB 176 AKAAAKKKATKSPAKPKAVPKPAKPKAAKPKAAKPKAAKAAKAAKAAK 224

OS Mycobacterium smegmatis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCB1\_taxonomy=1772;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 70084 / mc(2)155;  
 RX MEDLINE=99110209; PubMed=9894918;  
 RA Lee B.H., Murugaesu-Cei B., Dick T.;  
 RT "Upregulation of a histone-like protein in dormant Mycobacterium  
 RT smegmatis."; Mol. Genet. 260:475-479(1998).  
 CC -1- FUNCTION: Histone-like DNA-binding protein which is capable of  
 CC wrapping DNA to stabilize it, and thus to prevent its denaturation  
 CC under extreme environmental conditions (By similarity).  
 CC -1- SIMILARITY: Belongs to the bacterial histone-like protein family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF068138; AAD13809.1; -;  
 DR HSSP, P02346; IHU.  
 DR InterPro: IPR000119; Bac DNABind.  
 DR Pfam, PF00216; Bac DNA binding; 1.  
 DR PRINTS; PR01727; DNABINDINGHU.  
 DR ProDom; PD000945; Bac DNABind; 1.  
 DR SMART; SMO0411; BHL; 1.  
 DR PROSITE; PS00045; HISTONE LIKE; 1.  
 KM DNA-binding; DNA condensation; Repeat.  
 FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.  
 FT DOMAIN 101 205 DEGENERATE REPEATS REGION.  
 FT SEQUENCE 208 AA; 21230 MW; CASF577F61F7EP09 CRC64;  
 SQ  
 Query Match 30.8%; Score 160; DB 1; Length 208;  
 Best Local Similarity 54.3%; Pred. No. 0.00021;  
 Matches 57; Conservative 3; Mismatches 31; Indels 14; Gaps 5;  
 QY 1 AKTKAKAEKAYAKKAAKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAE 60  
 DB 111 AKTKAKKAP---AKKAAKKTATTAAKKAPAKKAA-----TKAPAKKAAATK-APAKKAA 161  
 QY 61 YKAEKKYAKKAAKAEKKEYYAAAEAKKAEAKKAYAKKAEAKKAAKXA 105  
 DB 162 TKAPKKAATKAPK-KKAAAKAPAKKA-----ATAPAKKAAAKKA 201

Search completed: April 20, 2004, 23:16:19  
 Job time : 15.8636 secs



```

RESULT 2
Q8WXL1 ID Q8WXL1 PRELIMINARY; PRT; 372 AA.
AC Q8WXL1;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Tola protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_Taxid=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96198174; PubMed=8626299;
RA Rodriguez-Herva J.J.; Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
lipoprotein (PAL) is involved in maintenance of the integrity of the
cell envelope."
RT J. Bacteriol. 178:1699-1706(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX Rodriguez-Herva J.J.;
RT Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96422022; PubMed=8824639;
RA Rodriguez-Herva J.J.; Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida."
RX EMBL; X74218; CAB50780.1; -.
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleosome; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; F:nucleosome assembly; IEA.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR006260; TONB_C.
DR PRINTS; PR00624; HISTONEH5.
DR TIGFAMS; TIGR01352; tonB_Cterm; 1.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 36.7%; Score 190.5; DB 2; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.3e-06;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAE-----AKKKAKAEAKKAYAKAA 54
DB 118 AKKAEEDA-AKAEAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 176
QY 55 K--AKKEAYKAEKAYAKAAEKKAYAAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 109
DB 177 KKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAK 233

RESULT 3
Q88N16 ID Q88N16 PRELIMINARY; PRT; 372 AA.
AC Q88N16;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Biopolymer transport protein TOLA.
GN TOLA OR PPI221.
OS Pseudomonas putida (strain KT2440).

```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_Taxid=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E.; Weinel C.; Paulsen I.T.; Dodson R.J.; Hilbert H.;
RA Martins dos Santos V.A.P.; Fouts D.E.; Gill S.R.; Pop M.; Holmes M.;
RA Brinkac L.; Beanan M.; DeBoy R.T.; Daugherty S.; Kolonay J.;
RA Madupu R.; Nelson W.; White O.; Peterson J.; Khouri H.; Hance I.;
RA Chris Lee P.; Holtzapple E.; Scanlan D.; Tran K.; Moazed A.;
RA Utterback T.; Rizzo M.; Lee K.; Kosack D.; Moesli D.; Medler H.;
RA Lauber J.; Stepanovic D.; Hoheisel J.; Stratzel M.; Helm S.;
RA Kiewitz C.; Eisen J.; Timmis K.N.; Duesterhoeft A.; Tsemmler B.;
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
metabolically versatile Pseudomonas putida KT2440."
RT Environ. Microbiol. 4:799-808(2002).
DR EMBL; AB016778; AAN66845.1; -.
DR TIGR; PPI221; -.
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006334; F:nucleosome assembly; IEA.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
KM Complete proteome.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785ECC3C0BC CRC64;

Query Match 36.7%; Score 190.5; DB 16; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.3e-06;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAE-----AKKKAKAEKAKKAKAA 54
DB 118 AKKAEEDA-AKAEAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAE 176
QY 55 K--AKKEAYKAEKAYAKAAEKKAYAAAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 109
DB 177 KKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAKKKAABEAKKAAEDAK 233

RESULT 4
Q8T5C8 ID Q8T5C8 PRELIMINARY; PRT; 1866 AA.
AC Q8T5C8;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE MaebL.
GN MAEBL.
OS Plasmodium vivax.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5855;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Salvador;
RA Michon P.; Stevens J.R.; Kaneko O.; Adams J.H.;
RT "Evolutionary relationships of conserved cytosine-rich motifs in
adhesive molecules of malaria parasites."
RT Mol. Biol. Evol. 0:0-0(2002).
DR EMBL; AY042083; AAL10508.1; -.
DR InterPro; IPR008602; Duffy_binding.
DR Pfam; PF05424; Duffy_binding; 1.
SQ SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;

Query Match 36.7%; Score 190.5; DB 5; Length 1866;
Best Local Similarity 49.6%; Pred. No. 4.1e-05;
Matches 61; Conservative 17; Mismatches 22; Indels 23; Gaps 6;

QY 5 AKKAEKAYAKKAKAK-----EKKAYAKKEAKAYKAAE-----AKKKAKAEKAKKAKAA 50
DB 1262 AKKAE--AKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAKKAEAK 1319

```

RA Waterston R., Wilson R.K.;

1. *What is the purpose of this study?*

DB 120 AEEAAKQAE LKQQAEEAAAKAADAKAKAEADDKA--AEAAKKAADAKKKAEAEAAK 177

[illegible]

RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / Biovar Orientalis;  
RX MEDLINE=21470413; PubMed=11586350;  
RA Parthill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahlia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feltwell T., Hanlin N., Holroyd S., Jagals K., Karlyshev A.V.,  
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,  
RL "Genome sequence of *Yersinia pestis*, the causative agent of plague.";  
RT Nature 413:523-527(2001).  
DR EMBL; AJ414146; CAC89966.1; -  
PIR; AC0138; AC0138.  
KW Complete proteome.  
SQ SEQUENCE 388 AA; 40424 MW; 81447B04B30A7E7C CRC64;

Query Match 34.3%; Score 178; DB 16; Length 388;  
Best Local Similarity 52.1%; Pred. No. 6,8e-05;  
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

QY 2 KKYAKKAEKAYAKKAKAKK-----AYAKKEA-KAYKA-AAAKKKAKAEKAYAKEA 53  
Db 140 KQAAEQQKIAAAVAAKAKEQKQETAQAQKAPADKIVKQQAQAQKAEAEKAEKAAVA 199  
QY 54 AAKKAEAYKAEKAYKAYAAK-----AEKKEYAAAEKKEAAK-AAKAEAAKAAKAA 105  
Db 200 AAARKQA-DADAKKAVVEAEKAAADAADAEKKAADAE-KKAAAKKVVAAAAKKKAAAE 257  
QY 106 A 106  
Db 258 A 258

RESULT 10  
Q8CZ28 PRELIMINARY; PRT; 393 AA.  
AC Q8CZ28;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Membrane spanning protein.  
GN TOLA OR Y3056.  
OS *Yersinia pestis*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; *Yersinia*.  
OC NCBI\_TaxID=632;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=KIM5 / Biovar Medievalis;  
RX MEDLINE=22137863; PubMed=12142430;  
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Lise P.,  
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,  
RA Straleny S.C., McDonough K.A., Niles M.L., Watson J.S., Blattner F.R.,  
RA Perry R.D.;  
RT "Genome sequence of *Yersinia pestis* KIM.";  
RL J. Bacteriol. 184:4601-4611(2002).  
DR EMBL; AB013906; AAM86606.1; -  
SQ SEQUENCE 393 AA; 41012 MW; 1E3E4F87E533481 CRC64;

Query Match 34.3%; Score 178; DB 16; Length 393;  
Best Local Similarity 52.1%; Pred. No. 6,9e-05;  
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

QY 2 KKYAKKAEKAYAKKAKAKK-----AYAKKEA-KAYKA-AAAKKKAKAEKAYAKEA 53  
Db 145 KQAAEQQKIAAAVAAKAKEQKQETAQAQKAPADKIVKQQAQAQKAEAEKAEKAAVA 204  
QY 54 AAKKAEAYKAEKAYKAYAAK-----AEKKEYAAAEKKEAAK-AAKAEAAKAAKAA 105  
Db 205 AAARKQA-DADAKKAVVEAEKAAADAADAEKKAADAE-KKAAAKKVVAAAAKKKAAAE 262











PA (TEVA-) TEVA PHARM USA INC.

PI Gad A, Lis D;

DR WPI; 2000-317499/27  
xx

PT Copolymer 1 related polypeptides used as molecular weight markers for  
PT glatiramer acetate and for treatment and prevention of immune diseases  
VV

PS Claim 10; Page 14; 72pp; English

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 104

CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present

DR WPI; 2000-317499/27

the invention describes polypeptides (1) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glutarimide acetate related terpolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune ophtholitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopenia purpura, colitis, contact sensitivity diseases, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glutarimide acetate molecules, which makes them ideal for use as molecular weight markers.

CC AAY8257 to AAH8257 represent specifically claimed copolymer molecular  
CC weight TV-marker polypeptides from the present invention. The present  
CC invention describes polypeptides (I) for determining the molecular weight  
CC of a copolymer (CP), which has an identified molecular weight and an  
CC amino acid composition corresponding to the copolymer. The polypeptides  
CC of the invention are used as molecular weight markers for glutarimer  
CC acetate related tetrapolymers. The polypeptides may also be used for  
CC treating and preventing immune diseases in a mammal. Autoimmune diseases  
CC which may be treated include either cell-mediated or antibody-mediated  
CC diseases. Such diseases include arthritic conditions, demyelinating  
CC diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid  
CC arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune  
CC celiacitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's  
CC disease, chronic immune thrombocytopenia purpura, colitis, contact

Sequence 77 AA

Query Match	55.7%;	Score 289;	DB 3;	Length 77;
Best Local Similarity	67.0%;	Pred. No. 3.1e-17;		
Matches 73; Conservative	1;	Mismatches 3;	Indels 32;	Gaps 4

mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical

QY 1 AKKVAKKAEEKVYAKKAKAAKEKKAYAKKAEAKAYTAAAEAKKKAKAEKKYAKKEAAKAKKEA 600  
||||| : ||| |||||  
Db 1 AKKYAKK-EKAYAKKA-----EKAQKAQAAKAYTAAEAQKKA----- 360

CC makes them ideal for use as molecular weight markers  
XX  
SQ Sequence 66 AA;

```
QY      61 YKAAKKYAKAKAKEKCYAAAEEAKKCAFAKAYKCEAAAKAAACGAAYYEA 109
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd      37 -KAAKCKYAKAACAEEKCYAAAAEAK-----YKBAKAAKAAAKGAAYEA  77
```

Query Match	44.0%;	Score 228.5;	DB 3;	length 66;
Best Local Similarity	56.9%;	Pred. No. 3e-12;		
Matches 62; Conservative	0;	Mismatches 4;	Indels 43;	Gaps 4

RESULT 4  
AAy82574  
ID AAy82574 standard; peptide; 66 AA.

Db 1 AKYAKK-EKAYAKAKA-----EAKAKA----- 25

AC	AAV82574;
XX	
DT	28-JUL-2000 (first entry)

Db 26 - KAAKCTAKAKAKEKEYPAAEA-----YKAAKAKAAKAAEAIEA 66

DE	Copolymer molecular weight	TV-marker amino acid sequence	SEQ ID NO:4
DE			

## RESULT 5

XX Copolymer; molecular weight marker; TV-marker; immune disease;  
KM glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;  
KM osteopontin; immunosuppressive; antithyroid; antiinflammatory;  
KM antidiabetic; thyromimetic; haemostatic; antiproliferic; dermatologic;  
KM antinaeumatic; immunosuppressive; demyelinating disease; osteoarthritis;  
KM inflammatory condition; multiple sclerosis; rheumatoid arthritis;  
KM Crohn's disease; chronic immune thrombocytopenia purpura; colitis;  
KM diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis  
KM Hashimoto's disease; idiopathic myxedema; myasthenia gravis;  
KM pemphigus vulgaris; systemic lupus erythematosus.

AAR06445  
 ID AAR06445 standard; protein; 154 AA.  
 XX  
 XX  
 AC  
 XX  
 AAR06445;  
 DT 25-MAR-2003 (revised)  
 DT 03-JAN-1991 (first entry)  
 XX  
 XX Recombinant copolymer 1-77, myelin basic protein analogue.  
 DE Recombinant copolymer 1; COP-1-77, myelin basic protein; MBP;  
 XX  
 XM

XX Unidentified.  
OS  
XX  
PN WO200018794-A1.

XX immunological activity; autoimmune encephalomyelitis; multiple sclerosis.  
XX  
XX synthetic.  
XX

PN EP363620-A.  
 XX  
 PD 22-AUG-1990.  
 XX  
 PF 16-FEB-1990; 90EP-00301700.  
 XX  
 PR 17-FEB-1989; 89US-00312541.  
 PR 07-FEB-1990; 90US-00473845.  
 XX  
 PA (REK ) REPLIGEN CORP.  
 XX  
 PI Cook KS;  
 XX  
 DR WPI; 1990-255848/34.  
 DR N-PSDB; AAQ05664.  
 XX  
 PT Producing genes encoding random polymers of aminoacid(s) - for producing  
 PT recombinant polypeptide(s) with biological and/or immunological activity.  
 XX  
 PS Disclosure; Fig 11; 25pp; English.

CC To improve the expression of rCOP-1 polypeptides in E. coli, genes coding  
 CC for rCOP-1-77 were subcloned from pREV 2.1 to PBG3-2delAN (deposit: 20-  
 CC NOV-1984 US4691009, NRL B-15910), a plasmid used to express Protein A.  
 CC The resulting plasmids encode fusion proteins consisting of beta-  
 CC glucuronidase, Protein A, and rCOP-1 sequences. A methionine residue  
 CC occurs between the Protein A and rCOP-1 sequences, originating from the  
 CC 5' linker sequence, in order that the COP-1 polypeptide may be cleaved  
 CC from the fusion protein. rCOP-1-77 contains oligonucleotide duplexes  
 CC including the following segments: YKK, EAK, NAK, AAK, and AAA. The N-  
 CC terminal alanine residue is left behind following CNBr cleavage of the  
 CC fusion protein. The product prevents or arrests experimental autoimmune  
 CC encephalomyelitis. They are used to prevent, arrest or control a  
 CC demyelinating disorder, e.g. multiple sclerosis. They may also be used as  
 CC additives to hair care products to confer beneficial effects on damaged  
 CC hair or as supplements for diets deficient in certain amino acids. See  
 CC also AAQ05665. (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 XX  
 SQ Sequence 154 AA;

Query Match 40.3%; Score 209; DB 2; Length 154;  
 Best Local Similarity 53.7%; Pred No. 3, 1e-10;  
 Matches 72; Conservative 8; Mismatches 20; Indels 34; Gaps 9;

QY 2 KKYAKAEKA-----YAKAKAAKKKKYAA-----KKKAAAYAAEA-----KKAK-A 44  
 DB 24 KKKAKAEKAKKAKYKKYKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 81  
 QY 45 EKKYAKAEKAAK-----KKEAYKAEKAKYAAKAAK-----KAYAAKAEKAAEA--K 91  
 DB 82 EYKKKAKAAAEYKKKEAEKAEYKKYKKKAKKAKYKKKAKAEKAAKAAKAAKAAKAAKAEYK 141  
 QY 92 AYKAAEAAVAAKAA 105  
 DB 142 KYKKEAEK--AKEA 153

RESULT 6  
 ABU40185  
 ID ABU40185 standard; protein; 372 AA.  
 XX  
 AC ABU40185;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #25712.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Pseudomonas putida.  
 XX  
 FN WO200277183-A2.  
 XX

PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002MO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362599P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Traxwick JD, Carr GJ, Yamamoto R, Foreysch RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA44055.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 68109; 176pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 CC  
 XX  
 SQ Sequence 372 AA;

Query Match 36.7%; Score 190.5; DB 6; Length 372;  
 Best Local Similarity 50.4%; Pred No. 2, 7e-08;  
 Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

QY 5 AKKAEKAYKKKAAK-----EKKAYAKKAAKAYKAAE-----AKKAAKAEKAYKAAEA 54  
 DB 118 AKKAEKAAEAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 176  
 QY 55 K--AKKAEKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 109  
 DB 177 KKAABEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAAEAKKAA 233

RESULT 7  
 ABU27824

ID ABU27824 standard; protein; 428 AA.  
 AC ABU27824;  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by Prokaryotic essential gene #13351.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Enterobacter cloacae.  
 PN W020027183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR N-PSDB; ACA31694.  
 DR WPI; 2003-029926/02.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 55748; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pat\_sequences  
 XX  
 XX Sequence 428 AA;  
 Query Match 36.6%; Score 190; DB 6; Length 428;

Best Local Similarity 50.8%; Pred. No. 3.4e-08;  
 Matches 64; Conservative 8; Mismatches 32; Indels 22; Gaps 5;  
 QY 1 AKKTA-----KAEKAYAKKAAKAEKAYAKKAAKAYAAEAKKKAKAA-----K 47  
 DB 167 AKKAAADQCKAEAEAKKAAADQCKAEAEAKKAA--AAADQCKAEAEAKKAAQAEK 224  
 QY 48 KAKAEAAKKAAYKAEKAYKAKAEKKEVAAAE-----KKAEAAK-----YKAEAKA 100  
 DB 225 KAAAEAAKKAAYKAEKAYKAKAEKKEVAAAE-----KKAEAAK-----YKAEAKA 284  
 QY 101 AAKKAA 106  
 DB 285 AAKKAA 290  
 RESULT 8  
 ID ABU28559  
 ID ABU28559 standard; protein; 421 AA.  
 AC ABU28559;  
 DT 19-JUN-2003 (first entry)  
 DE Protein encoded by Prokaryotic essential gene #14086.  
 KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 OS Escherichia coli.  
 PN W020027183-A2.  
 PD 03-OCT-2002.  
 PF 21-MAR-2002; 2002WO-US009107.  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR N-PSDB; ACA32429.  
 DR WPI; 2003-029926/02.  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 56483; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

















```

; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68109
;
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Pseudomonas putida
;
; US-10-282-122A-68109
;
Query Match      36.7%; Score 190.5; DB 12; Length 372;
Best Local Similarity 50.4%; Pred. No. 9.1e-08;
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4;

Qy 5 AKKAERKAYAKKAKKAAK-----EKKAAYAKKKAAYKAAE-----AKKKAKAERKAYAKAA 54
Db 118 AKKAEDA-AKKAABAERKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 176
Qy 55 K-AKKAERKAYAKKAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 109
Db 177 KKAABAERKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 233

RESULT 6
US-10-282-122A-55748
; Sequence 55748, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```

; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55748
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
;
; US-10-282-122A-55748
;
Query Match      36.6%; Score 190; DB 12; Length 428;
Best Local Similarity 50.8%; Pred. No. 1.2e-07;
Matches 64; Conservative 8; Mismatches 32; Indels 22; Gaps 5;

Qy 1 AKKA-----KKAERKAYAKKAKKAYAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 47
Db 167 AKKAADAKKAAERKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 224
Qy 48 KYAKEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 100
Db 225 KAAAEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 284
Qy 101 AKKAA 106
Db 285 AKKAA 290

RESULT 7
US-10-282-122A-56483
; Sequence 56483, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 56483
; LENGTH: 421
; TYPE: PRF
; ORGANISM: Escherichia coli
US-10-282-122A-56483

Query Match          36.5%; Score 189.5; DB 12; Length 421;
Best Local Similarity 53.0%; Pred. No. 1,2e-07;
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5

QY      1 AKTYAKAAE---KAYAKKAAKAKKAYAKKEKAYAYAAEAKKKAANEKTYAK-EAAK 55
      :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      120 AEEAAACQAEIKQKQKEAAAKAAADAAKKAABADKA--AEEAAKAAADAKKQAEAEAAK 177
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      56 AKKEAY-KAEAKKYAAAEAEKKEEYAAAEAAK---AEAAKYVYAEAAKAAAKAA 106
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      178 AAEEAQKQAEAAALAKKGAEEAAAEAAAEAAKKAATAEAEAAAEAEKKAANAEEKA 232
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```

RESULT 8
US-10-282-122A-75047
Sequence 75047, Application US/10282122A
Publication No. US20040029122A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyckind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347
PRIORITY FILING DATE: 2000-09-09
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/267,636
PRIORITY FILING DATE: 2001-02-09
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
Remaining Priority Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 75047
LENGTH: 407
TYPE: PRT
ORGANISM: Salmonella typhimurium
US-10-282-122A-75047

```

Query Match	35.5%;	Score 184.5;	DB 12;	Length 407;
Best Local Similarity	55.5%;	Pred. No. 3e-07;		
Matches 66;	Conservative 11;	Mismatches 25;	Indels 17;	Gaps 8

[illegible]

```

RESULT 9
US-10-282-122A-59321
; Sequence 59321, Application US/10282122A
; Publication No. US60040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EPIITA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59321
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59321

```

```

Query Match      35.0%; Score 181.5; DB 12; Length 323;
Local Similarity 52.7%; Pred. No. 4,1e-07;
Matches          59; Conservative    8; Mismatches   34; Indels   11; Gaps    4

QY  1 AKKAAKKAKKAAKKAAYAKEKAYAAAEKKAQAEEKTAKEAK-----55
DB  47 ADRQAQAQQAAARADPAK--KQEA---AARAAEAKQAQAEEAKKAQEKAQAEA 102
QY  56 -AKGEYAQAATVAAAAREKEVAEAEEKGAQAAYAAEAQAQAEEA 106
DB  103 AAQKQAQEEKKAQAQEAQAQAQAEE--KAEEKAQAQAQAEEKKAQAEEA 153

```

## RESULT 10

	Query Match	34.8%	Score 180.5;	DB 9;	Length 56;
	Best Local Similarity	45.9%	Pred. No. 7.9e-08;		
	Matches	50;	Conservative	3;	Mismatches 3; Indels 53; Gaps 3;
QY	1 AKTAAKAEKYAKGAKAAREKKAYAKKEAGVYYOALENKKAKADAKKYATEBAAKAKEEA	60			
Dd	1 AKTAAK-K-EKVAYAKKA-----EKAKAKAEALAVYPAAEKKPKAEK-	39			
Qy	61 YKAETKTAVAAAARAKEKETAAAAEAKCAAPAAATVTAEEAAKAAAKAEAYEA	109			
bB	40 -----YAAEAAKAAKAAEAAYA	56			

[illegible]

	Query Match	34.5%	Score 179	DB 14	Length 347
Best Local Similarity	43.1%	Prod. No. 6,9e-07			
Matches	54	Conservative 17	Mismatches 29	Indels 10	Gaps 4
QY	6	KKAEEAAYAKKAAYEK-----KAAKKEAKAYKAAEAKQAKAAKAYKAAEAAKAKE	59		
DB	99	QLEEQOYAAAKAAEOKKADAEARQAEOKAAEAKKADAEKKAEEAQAQAKOKROADIANKR	158		
QY	60	AYKAAKAYTAAKAAAEKKEYYAAAEAKKAAEAAKAYKAAEAAKAAEAAKAAEAA	109		
QY	159	A-EDAAK--KAAEDAKKK--AAEDAKKGALEAAEKKAAAEAAAKKAAAYA	204		

RESULT 13  
US-10-051-643-201  
; Sequence 201, Application US/10051643  
; Publication No. US20020197265A1  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James P.

```

; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; TITLE OF INVENTION: System using Mycobacterium Vaccae
; FILE REFERENCE: 11000.1008c2
; CURRENT APPLICATION NUMBER: US/10/051,643
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 201
; LENGTH: 223
; TYPE: PRN
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

```

```

Query Match      34.3%; Score 178; DB 13; Length 223;
Best Local Similarity 54.5%; Pred. No. 5, 2e-07;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;

```

```

QY 1 AKKYAKK--AEKAYAKKAAKE--KAYAKKAYKAAKKAKEAKKAYAKKAAK 55
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112 ARKAARAKPAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 56 AKKEAYKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 AKKAPAKKAAATKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-10-205-979-52
; Sequence 52, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Abernethy, Nevlin
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; TITLE OF INVENTION: of Immune Responses
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 52
; LENGTH: 223
; TYPE: PRN
; ORGANISM: Mycobacterium vaccae
US-10-205-979-52

```

```

Query Match      34.3%; Score 178; DB 14; Length 223;
Best Local Similarity 54.5%; Pred. No. 5, 2e-07;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;
QY 1 AKKYAKK--AEKAYAKKAAKE--KAYAKKAYKAAKKAKEAKKAYAKKAAK 55
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112 ARKAARAKPAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 56 AKKEAYKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 171 AKKAPAKKAAATKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
US-10-282-122A-75772
; Sequence 75772, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/151,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 75772
; LENGTH: 376
; TYPE: PRN
; ORGANISM: Salmonella typhi
US-10-282-122A-75772

```

```

Query Match      34.3%; Score 178; DB 12; Length 376;
Best Local Similarity 52.3%; Pred. No. 9e-07;
Matches 58; Conservative 10; Mismatches 33; Indels 10; Gaps 4;
QY 1 AKKYAKKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 137 AKAAADAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 YKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 191 -KAEAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Search completed: April 20, 2004, 23:27:18
Job time : 54.7922 secs

```













